NOVARTIS

Clinical Development and Medical Affairs

CACZ885M2301/canakinumab / NCT01327846

# Canakinumab Anti-inflammatory Thrombosis Outcomes Study (CANTOS) - A randomized, double-blind, placebo-controlled, event driven trial of quarterly subcutaneous canakinumab in the prevention of recurrent cardiovascular events among stable post-myocardial infarction patients with elevated hsCRP

## Detailed Statistical Methodology

Author(s): [redacted]

Document type: Analysis Plan Documentation

Document status: Final; Amendment 4

Release date: 14 April 2017

Number of pages: 89

## Table of contents

## List of tables



## List of figures

## List of acronyms

| | |
|---|---|
| AE | Adverse event |
| AHD | Antihyperglycemic drug |
| ALP | Alkaline phosphatase |
| ALT | Alanine aminotransferase (SGPT) |
| AST | Aspartate aminotransferase |
| ATC | Anatomical Therapeutic Chemical classification system |
| BCa | Bias-corrected and accelerated bootstrap |
| BMI | Body Mass Index |
| BNP | Brain Natriuretic Peptide |
| [redacted] | [redacted] |
| CABG | Coronary Artery Bypass Graft |
| CAD | Coronary artery disease |
| CANTOS | Canakinumab Anti-inflammatory Thrombosis Outcomes Study |
| CCU | critical care unit |
| CDBL | Clinical Database Log |
| CEC | Clinical Events Committee |
| CHD | Coronary Heart Disease |
| COPD | Chronic Obstructive Pulmonary Disease |
| CREDI | Novartis Clinical REsearch Documentation and Information system |
| CRF | Case Report Form |
| CRS | Case Retrieval Strategy |
| CV | Cardiovascular |
| DBL | Conjugated (direct) bilirubin |
| DBP | Diastolic Blood Pressure |
| DHP | Data Handling Plan |
| DILI | Drug-induced liver injury |
| DMC | Data Monitoring Committee |
| ECG | Electrocardiogram |
| eGFR | estimated Glomerular Filtration Rate |
| FAS | Full Analysis Set |
| FPG | Fasting Plasma Glucose |
| HbA1c | Glycosylated hemoglobin |
| HDL | High Density Lipoprotein |
| HR | Heart Rate |
| hsCRP | high sensitivity C-reactive protein |
| ICU | intensive care unit |
| IDL | Intermediate density lipoprotein |
| Il-1β / Il-6 | Interleukin-1β, Interleukin-6 |
| IRT | Interactive Response Technology |
| IP | Investigational Product |
| IVRS | Interactive Voice Response System |
| LAD | Left Anterior Descending artery |
| LCX | left circumflex artery |
| LDL | Low Density Lipoprotein |
| LFT | Liver Function Test |
| LLN | Lower Limit of the Normal range |
| LPLV | Last Patient Last Visit |
| MACE | Major Adverse Cardiovascular Events (CV death, MI or stroke) |
| MAR | Missing At Random |

| | |
|---|---|
| MCMC | Markov chain Monte Carlo |
| MedDRA | Medical Dictionary for Regulatory Activities |
| MI | Myocardial infarction |
| NMQ | Novartis MedDRA Query |
| NOD | New Onset Diabetes |
| NovDTD | Novartis Drug/Therapy Dictionary |
| NT-proBNP | N-terminal pro-Brain Natriuretic Peptide |
| OC/RDC | Novartis Oracle Clinical/Remote Data Capture system |
| PCI | Percutaneous Coronary Intervention |
| PD | Protocol Deviation or Pharmacodynamics |
| [redacted] | [redacted] |
| PPS | Per-Protocol Set |
| PT | MedDRA Preferred Term |
| QTc | heart rate-corrected QT interval |
| QTcB | QTc (Bazett's correction) |
| QTcF | QTc (Fridericia's correction) |
| RAAS | Renin-angiotensin-aldosterone system |
| RCA | Right coronary artery |
| RMP | Novartis Risk Management Plan |
| SAE | Serious Adverse Event |
| SAF | Safety Set |
| SBP | Systolic Blood Pressure |
| SMQ | Standardized MedDRA query |
| SOC | (primary) System Organ Class |
| SPP | Novartis Safety Profiling Plan |
| T1DM | Type 1 Diabetes Mellitus |
| T2DM | Type 2 Diabetes Mellitus |
| TBL | Total bilirubin |
| TIA | Transient Ischemic Attack |
| TNF | Tumor necrosis factors |
| [redacted] | [redacted] |
| VLDL | Very low-density lipoprotein |

Document History – Changes compared to previous versions of RAP module 3

**Amendment 4, April 14, 2017**

| Section | Changes |
|---|---|
| 2.1 | Table 2-1: Addition of PD for site closure due to GCP non-compliance |
| 2.2 | Table 2.3-1 Age subgroup categories clarified to make clear that these are two separate subgroup variables |
| 2.2 | Table 2.3-1: Added subgroup as requested by FDA (presence of MI prior to index MI – yes/no) |
| 2.2 | Table 2.3-2 Change made Age subgroup categories to clarify this is two separate subgroup variables (consistent with Table 2.3-1) |
| 2.3.1 | Added text to highlight different terminology used in the protocol |
| 2.3.8 | Amended to reflect that the LPLV date will be used as the cut-off date since all events are planned to be adjudicated up to the last patients' visit |
| 5 | Removed Table 5-1. Revised text referencing table 5-1 to state that summaries should be based on the protocol visit schedule. Updated references to Table 5-2 due to table re-numbering. |

**Amendment 3, July 19, 2016**

| Section | Changes |
|---|---|
| Various | Various changes to text are made to clarify the intent. Minor changes are not reflected below. Changes that result in a change from what was originally intended are explained below. |
| 1 | Changed wording to reflect that the data from the washout and extension phases of the study will be addressed in a separate analysis plan and study report. |
| 2.2 | Table 2.3-3: removed subgroup related to patients entering washout who had stopped medication prior to final main trial assessment. |
| 2.3.8 | Clarifying that the analysis cutoff date could be earlier than the LPLV. In this case, events will only be considered as investigator reported. |
| 2.3.9 | - Removed planned sensitivity analyses for examining the assumption of non-informative censoring. This will be investigated, but not generally included in the study report.<br>- Clarified that censoring date may not be the same for all time to event endpoints. |
| 2.4 | Figure 2-3 updated to match corresponding figure in study protocol. |
| 5 | Definitions of medication categories updated. |
| 6 | Removing the text about "post treatment phase" as it was resulting in confusion as to what was intended. Clarified main analyses will be based events occurring in the double blind phase. |
| 6.1.2 | Updated censoring algorithm to align with use of the cutoff date. |
| 6.3.1 | - Covariate adjusted analysis of primary endpoint changed to "may be considered." This is an exploratory analysis and not planned to be part of the study report.<br>- Added that the association between hsCRP changes and outcomes will be explored, this is already mentioned in the current protocol. |
| 6.3.2 | - Added supportive analyses for diabetes related time to event endpoints.<br>- Removed reference to analysis of washout period data.<br>- Added statement that washout period data will be analyzed in a separate study |

| Section | Changes |
|---|---|
| | report or an addendum to the study report. |
| [redacted] | [redacted] |
| 6.5 | Dose response model approach revised to reflect current plan for dose response modelling. |
| 7 | - Removed text related to summarizing data by trial part to avoid confusion between the main AE tables, which are pooled across trial parts and some specific analyses that present data separately by trial part.<br>- Revised categories used for by-time-period analyses. |
| 7.3.1 | - Clarified on-treatment definition for AEs.<br>- Removed plan for time-to-event analysis or AEs. |
| 7.3.2 | - Updated notable criteria in Table 7-1.<br>- Added detailed description for analysis of long term effects on kidney function. |
| 7.4 | Updated list of adverse events of special interest to align with current safety profiling plan. |
| 7.4.1 | - Added by-time-period analyses for infections to align with other adverse event analyses.<br>- Model for analysis of infections is revised to align with model for other adverse event analyses.<br>- Analysis of duration of infections revised to only present a descriptive summary. Negative binomial model analysis is removed. |
| 7.4.2 | Removed ordinal regression analysis of LFT parameters; results are summarized using frequency tables. |
| 7.4.3 | Additional summary tables added to align with other adverse event analyses. |
| 7.4.4 | Removed summary of early drops in platelet counts. |
| 7.4.8 – 7.4.12 | - Topic of long term effects on kidney function was moved into Section 7.4.12 (Other safety topics). Analysis description is covered in section 7.3.2. (Laboratory data).<br>- Topic of QT prolongation was moved into section 7.4.12 (Other Safety topics) and will be analyzed under section 7.3.4 (ECG).<br>- Section on Hypoglycemia reordered to section 7.4.8.<br>- Added new sections 7.4.9-7.4.11 (Autoimmunity reactions; Injection site reactions; Vertigo/dizziness).<br>- Section on “Other safety topics” reordred to section 7.4.12. Section title changed to clarify that this section addresses other adverse events that are not part of the adverse events of special interest. |
| 10 | Removed; this will be addressed in a separate analysis plan and report. |
| 11 | Removed; this will be addressed in a separate analysis plan and report. |
| Appendix 1 | Removed; this will be addressed in a separate analysis plan and report. |
| Appendix 2 | Renumbered to Appendix 1. |

**Amendment 2, April 13, 2016**

| Section | Changes |
|---|---|
| Throughout | Editorial changes. |
| 2.1 (Table 2-2) | Updated Per protocol criteria. |
| 2.2 (Table 2.3-1) | - Detail added on the region subgroups. |

| Section | Changes |
|---|---|
| | - Clarified analyses would be done in all tables by region subgroup.<br>- Updated definition of Pre-diabetic and Diabetic. |
| 2.2 (Table 2.3-3) | - Details added to basis for diabetes diagnosis subgroups<br>- Baseline HbA1c/FPG categories subgroups updated. |
| 2.3.4 | On-treatment definition added. |
| 2.3.9 (previously 2.3.8) | - Added missing date imputation rules.<br>- Amended time to event section.<br>- MACE endpoint (and all individual components) censoring rules updated. |
| 3.1 | Changed the calculation of patient's age to be from date of start of screening instead of date of informed consent. |
| 3.2 | Updated sub-level categories. |
| 5 | - Added PCSK9 inhibitors to the list of non-statins.<br>- Added SGLT2 inhibitors to the list of other oral hypoglycemic agents.<br>- Added rules for the summary of statin use for the determination of the daily dose for each patient. |
| 6.3.2 | - Updated the definition of new onset diabetes.<br>- Updated the second set of criteria for determining new onset diabetes. |
| [redacted] | [redacted] |
| [redacted] | [redacted] |
| [redacted] | [redacted] |
| 7.3.1 | Amended the cut off for common events from 1% to 2%. |
| 7.3.3 (Table 7-2) | Amended clinically notable changes in vital signs criteria. |
| 7.4 | Inserted table 7-3 with an updated list of AEs of special interest. |
| 7.4.2 | - Clarified baseline glycemic and abnormal LFT subgroup analyses.<br>- Updated the covariates in the ordinal regression. |
| 7.4.3 | Amended the identification of malignancy. |
| 12 | Added references. |

**Amendment 1, June 26, 2014**

| Section | Changes |
|---|---|
| 2.2 | Updates to the per-protocol analysis sets. |
| 2.2 | Updates to the subgroup definitions. |
| 2.3.2 | Clarification of the hsCRP the baseline value. |
| 3 | Minor updates and clarifications to patient disposition, demographics, and medical history. |
| 6.1.1 | New section 6.1.1 was added for endpoint adjudication process. |
| 6.3.1 | - Addition of an additional early futility analysis to be performed at approximately 30% of the target number of 1400 patients have experienced CEC confirmed MACE, as per protocol amendment 8.<br>- Clarification that in case one active arm is stopped, the target number of patients with a MACE that is still to be collected has to be adjusted. For patients whose treatment arm is stopped per DMC recommendation, follow-up for cardiovascular (CV) and safety events will continue. |

| Section | Changes |
|---|---|
| | - A sensitivity analysis will be performed by performing the primary analysis excluding all patients who were unblinded during the double blind phase of the study.<br>- Change from baseline to the last hsCRP prior to the 1st confirmed MACE will be summarized by treatment. |
| 6.3.2 | - Clarification added that all identified cases of new onset of diabetes will be confirmed by the adjudication committee.<br>- Clarifications on subgroup analyses for the key secondary new onset of diabetes endpoint. |
| [redacted] | [redacted] |
| [redacted] | [redacted] |
| 7 | Several minor updates and clarifications. |
| 8 | Protocol amendment 8 introduced an additional early futility analysis to be performed at approximately 30% of the target number of 1400 patients have experienced CEC confirmed MACE. |
| 12 | Added new reference. |

# 1 Introduction

The Canakinumab Anti-inflammatory Thrombosis Outcomes Study (CANTOS; Novartis study code CACZ885M2301) is a randomized, double-blind, placebo-controlled, event-driven trial of quarterly subcutaneous canakinumab in the prevention of recurrent cardiovascular events among stable post-myocardial infarction patients with elevated hsCRP.

This document outlines the agreements for the Statistical Analysis Plan for CANTOS based on the discussions between [redacted] and Novartis. The detailed statistical analysis plan including additional supportive/[redacted], table shells and data set specifications will be prepared on this basis.

It is planned that after the end of the core study visit patients will be followed for safety in an extension study. Data collected after the end of the core study visit will not be part of the main CANTOS clinical database and analyses using data from this extension study will not be specified in this document.

# 2 Definitions

## 2.1 Analysis sets

The following analysis sets will be defined for statistical analysis:

- **Screened set** – All patients who signed the informed consent.
- **Randomized set** – All patients who received a randomization number, regardless of receiving trial medication.
- **Safety set (SAF)** – All patients who received at least one dose of study drug and have at least one post-baseline safety assessment. Of note, the statement that a patient had no adverse events also constitutes a safety assessment. Patients will be analyzed according to treatment received. Treatment received will be considered identical to the randomized treatment if the patient has during at least one visit received the two injections constituting the treatment assignment at randomization.

  For patients who cannot be assigned in this manner, the most frequent dose at an individual visit (as defined in table 2-1) will be considered the treatment received. In the case of more than one equally frequent dose (ties), the highest most frequent dose will be assigned.

  Note: Any patient who did not receive any injection of study drug during the study is automatically excluded from the safety set, due to the lack of at least one dose of study medication.

**Table 2-1 Definition of dose received at a visit**

| Sum of injection doses* [mg] | Dose at the visit |
|---|---|
| 0 (meaning ≥ 1 placebo injections, but not even a partial canakinumab injection) | Placebo |

| Sum of injection doses* [mg] | Dose at the visit |
|---|---|
| > 0 to ≤ 50 mg of canakinumab | Canakinumab 50 mg |
| > 50 to ≤ 150 mg of canakinumab | Canakinumab 150 mg |
| > 150 of canakinumab | Canakinumab 300 mg |

* Placebo injections counted as 0 mg

- **Full analysis set (FAS)** – All randomized patients. This is the primary efficacy population applied in all efficacy variables. Following the intent-to-treat principle, patients are analyzed according to the treatment they have been assigned to at the randomization. However, patients who have not been qualified for randomization and who have been inadvertently randomized into the study are excluded from FAS, provided these patients have not received study drug. (Note: The last part of the definition of the FAS is what is often referred to as misrandomized patients; i.e. patients for whom IVRS calls were made by the site either prematurely or inappropriately prior to confirmation of the patient's final randomization eligibility and double-blind medication was not administered to the patient. These patients would subsequently not continue to take part in the study or be followed-up. Misrandomized patients will not be included in the FAS, but they will be included in the Randomized Set. Further exclusions from the FAS may only be justified in exceptional circumstances; e.g., serious Good Clinical Practice violations).

  **Per protocol set 1 (PPS1)** – a subset of the FAS, consists of all randomized patients in FAS who take at least one dose of study medication and have no major protocol deviations affecting the primary analyses. Major protocol deviations leading to exclusion from PPS will be specified prior to database lock on a blinded basis and documented in a separate document. Data from patients discontinuing treatment will be censored a quarter year + 28 days after their last study drug injection.

All cases of prospectively defined protocol deviations will be identified prior to clinical database lock / unblinding and entered into a dedicated data panel as part of the locked database. Certain deviations may stipulate that only data up to the time of infraction will be included in the PP analysis, i.e. for analyses of the PP set data beyond this time-point will be ignored. This supplemental efficacy population is used to assess robustness of the primary analysis results. All exceptional cases and problems and the final decisions on the allocation of patients to populations will be fully defined and documented before data base lock (in particular before breaking the blind where applicable) and will be fully identified and summarized in the clinical study report as per the harmonized International Conference on Harmonization of technical requirements for registration of pharmaceuticals for human use tripartite guideline on statistical principles for clinical trials.

Criteria defining protocol deviations are provided in the Data Handling Plan (DHP). Protocol deviations will not lead to patient withdrawal from the study unless they indicate a significant risk to the patient's safety.

For the purpose of summarizing major protocol deviations, these will be considered to be those protocol deviations that lead to the exclusion from an analysis set. The current draft list of criteria leading to exclusions from analysis sets is shown in Table 2.2. Other currently

foreseen protocol deviations will be reported in the clinical study report, but these will not lead to an exclusion from any analysis set. It is not currently foreseen that any criteria would lead to an exclusion from PPS from date of deviation.

**Table 2.2 Criteria leading to exclusions from analysis sets**

| Analysis set | Criteria that cause a screened patient to be excluded |
|---|---|
| Screened | none |
| Safety (SAF) | Patients not receiving any dose of double-blind study medication |
| Randomized | Patient without a randomization number |
| | Patient was not randomized but took study drug (covered by the above) |
| Full Analysis Set (FAS) | Patient mis-randomized |
| | Patient without a randomization number |
| | Patient was not randomized but took study drug |
| | Site closed for GCP non-compliance (data integrity was impacted) |
| Per Protocol set 1 (PPS1) | Patient mis-randomized |
| | Patient without a randomization number |
| | Patient was not randomized but took study drug |
| | Patients not receiving any dose of double-blind study medication |
| | Patient participates in any other investigational drug or device study or patient has received an investigational drug or device <= 30days of Visit 1 except for previous DES trial on countries with approved DES devices |
| | Patient received study drug, and was later discontinued due to not meeting the entry criteria |
| | Major surgical procedure including PCI or CABG performed during the study but planned prior to randomization. |
| | CABG < 3 years prior to visit 1 |
| | Patient is currently (i.e. at randomization) taking a biological drug targeting the immune system (ie: TNF blockers, anakinra, rituximab, abatacept, tocilizumab) |
| | Documented MI does not meet entry criteria or Date of most recent MI = blank but there is a visit 2 |
| | Screening hsCRP < 2 mg/L or = blank |
| | Qualifying hsCRP measurement date < 28 days after date of prior PCI or Qualifying hsCRP measurement date < 28 days after date of qualifying MI |
| Exclusion of data from PPS from date onwards | Data from patients discontinuing treatment will be censored a quarter year + 28 days after their last study drug injection. |

## 2.2 Subgroup definitions

Generally subgroup analyses within a trial should be kept to a minimum. However, subgroup analyses need to be presented in submission documents like Summary of Clinical Efficacy, Summary of Clinical Safety, and Risk Management Plan and the clinical study report for the CANTOS trial will likely at least fulfill the first two of these roles.

The objective of the subgroup analyses is to evaluate the consistency of treatment effects across a wide variety of patient groups.

The primary composite (MACE) endpoint of CV death, nonfatal MI and nonfatal stroke, the key secondary variables and potentially other important efficacy variables will be analyzed for the subgroups shown in the Table 2.3-1.

**Table 2.3-1 Subgroup analyses for the primary major adverse cardiovascular event (MACE) endpoint and other key efficacy outcomes**

| Subgroups | Definition |
|---|---|
| Age (at study start) | <65, ≥65 years ;<br><75, ≥75 years |
| Sex | male, female |
| Race | Asian, Black, Caucasian, Other |
| Ethnicity | Hispanic / Latino, Asian, Other, Unknown |
| Type of qualifying MI | STEMI, non-STEMI, unknown |
| BMI at baseline | < 25, ≥ 25 & < 30, ≥ 30 kg/m$^2$ ; < 35, ≥ 35 kg/m$^2$ |
| Region | • Asia: China, India, Japan, South Korea, Taiwan<br>• Western Europe: Austria, Belgium, Germany, Greece, Iceland, Italy, Netherlands, Norway, Sweden, United Kingdom<br>• Central Europe: Bulgaria, Croatia, Czech Republic, Estonia, , Hungary, Latvia, Lithuania, Poland, Romania, Russian Federation, Serbia, Slovakia, Slovenia, Turkey<br>• Latin America: Argentina, Brazil, Colombia, Guatemala, Mexico, Peru<br>• North America: Canada, United States<br>• Others: Australia, South Africa<br><br>In addition to the above regions in all tables by region subgroup, an analysis for Japan (and Japan plus Korea) will also be presented separate regardless the size of enrollment in these countries to meet Japanese regulatory requests. |
| Glycemic status at baseline | Normoglycemic, pre-diabetic, diabetic<br>Pre-diabetic is defined by one of the followings<br>• HbA1c ≥5.7% & <6.5%<br>• FPG ≥100 & <126 mg/dl<br>• FPG ≥ 126 mg/dl at either Visit 1 or Visit 2 (or any unscheduled visit between, but only occurring on one visit)<br>• HbA1c ≥ 6.5% at either Visit 1 or Visit 2 (or any unscheduled visit between, but only occurring on one visit)<br><br>Diabetic is defined by one of the followings<br>• FPG ≥ 126 mg/dl at Visit 1 and Visit 2 (or any unscheduled visit between) |

| Subgroups | Definition |
|---|---|
| | • HbA1c ≥ 6.5% at Visit 1 and Visit 2 (or any unscheduled visit between)<br>• Combination of FPG ≥ 126 mg/dl and HbA1c ≥ 6.5% as confirmed at Visit 1 and Visit 2 (or any unscheduled visit between)<br>• Medical history of Type I or Type II diabetes mellitus<br>• Anti-diabetic medication use |
| Smoking status at baseline | Never, current, former |
| Baseline hsCRP level | ≤ 4, > 4 mg/L |
| Baseline LDL-C level | < 70 mg/dL, >= 70 mg/dL; Tertiles of baseline LDL-C (This will be done using the calculated LDL-C unless triglycerides > 400 mg/dL in which case a direct (i.e. measured) LDL-C will be used) |
| Baseline SBP level | < 130 mmHg, ≥ 130 mmHg |
| Baseline DBP level | < 80 mmHg, ≥ 80 mmHg |
| Baseline statin dose level | No, low, moderate, and high |
| Baseline aspirin usage/dose | Yes / No |
| Medical history of gout | Yes / No |
| Cardiovascular risk factors | |
| Hypertension | Yes / No |
| Dyslipidemia | Yes / No |
| Prior PCI | Yes / No |
| Prior CABG | Yes / No |
| Prior TIA/stroke | Yes / No |
| History of heart failure | Chronic Heart Failure occurrence (yes/no) |
| eGFR MDRD | < 60, ≥ 60 & < 90, ≥ 90 mL/min/1.73m$^2$ |
| Time since index MI | < 6 vs. ≥ 6 months ; < 12 vs. ≥ 12 months |
| Presence of MI prior to index MI | Yes / No |

A set of standard safety subgroups shown in Table 2.3-2 will by default be looked at for selected safety analyses (see section 7).

**Table 2.3-2 Standard safety subgroups**

| Subgroup | Definition |
|---|---|
| Age | <65, ≥65 years;<br><75, ≥75 years |
| Sex | male, female |

| Subgroup | Definition |
|---|---|
| Race | The same as above |
| Ethnicity | The same as above |
| Region | The same as above |
| Time since index MI | < 6 months vs. ≥ 6 months |
| BMI | The same as above |
| Medical history of gout | Yes / No |
| Co-existing medical conditions of interest: T2DM | Yes / No |
| Baseline hsCRP level | The same as above |

In addition to the two preceding sets of subgroups those specified in Table 2.3-3 will be used in some analyses as stated in the relevant analysis plan sections. They are shown here irrespective of what kind of analysis they will be used for to highlight that they needed to be included in analysis datasets.

**Table 2.3-3 Additional subgroups needed only for specific analyses**

| Subgroups | |
|---|---|
| Duration of diabetes | < 3, ≥ 3 & < 10, ≥ 10 years |
| Basis for diabetes diagnosis | • Medical History of Type I or Type II diabetes mellitus<br>• Medical History and Anti-diabetic medication<br>• Medical History and Baseline HbA1c/FPG<br>• Medical History and Anti-diabetic medication and Baseline HbA1c/FPG<br>• Anti-diabetic medication<br>• Anti-diabetic medication and Baseline HbA1c/FPG<br>• Baseline HbA1c/FPG |
| Baseline HbA1c categories | • HbA1c < 7% at randomization for deterioration of glycemic control analyses (HbA1c ≥ 7.5% at 6 month visit or beyond, confirmed by a second/unscheduled HbA1c measurement within 6 weeks) [i.e. Among those randomization diabetic patients, we look at the sub-group of patients with HbA1c < 7% at visit 2. Then we would analyze how many of those patients experience a deterioration of glycemic control as documented by an increase of HbA1c to >= 7.5% at 6 months or beyond, confirmed within 6 weeks.]<br>• For HbA1c benefit analysis in diabetics (HbA1c < 8%, ≥ 8%) [i.e. analysis of mean HbA1c change from baseline in T2D patients with baseline HbA1c < 8% and >= 8%.] |
| History of macular degeneration | Medical history or use of relevant medications |
| Baseline neutrophil levels | Abs neutrophils < 1.5 , ≥ 1.5 x $10^9$ /L<br>. |
| Baseline use of antithrombotic medications | Yes/No |
| key drugs for the trial population affected by CYP450 with a narrow therapeutic index (e.g. warfarin and amiodarone) | Yes/No<br>• |
| Vaccines | Yes/No |
| Medical history of gout, rheumatoid arthritis or other inflammatory diseases such as lupus or psoriatic arthritis | Yes/No |
| History of malignancy | Yes/No - Yes will correspond to patients with SMQ “Malignant or unspecified tumors” in Medical history |
| Baseline QTcF | Elevated or normal |

| Subgroups | |
|---|---|
| | Normal<br>• ≤ 430 msec for men, ≤ 450 msec for women<br>Elevated<br>• > 430 msec for men, > 450 msec for women. |
| Baseline heart rate | < 55 bpm versus ≥ 55 bpm |
| Alcohol history | Yes / No |
| Baseline platelet counts | < 150, ≥ 150 x $10^9$ /L |
| Early drops in platelet counts (or thrombocytopenia) Ordinal | < 25, ≥ 25 - <50, ≥ 50 - <75, ≥ 75 - < LLN, >=LLN at visit 3. Repeat at visit 4; |
| Early drops in platelet counts (or thrombocytopenia) defined by % change from baseline | < 20% and ≥ 20% drop from baseline to visit 3. Repeat at visit 4 |
| Early drops in neutrophils (or neutropenia) | <0.5, ≥0.5 - <1.0, ≥1.0 - <1.5, ≥1.5 - <LLN at visit 3. Repeat at visit 4 |
| Early drops in neutrophils (or neutropenia) defined by % change from baseline | < 20% and ≥ 20% drop from baseline to visit 3. Repeat at visit 4 |
| baseline neutropenia | CTC >=3 (yes/no) |

To support regional submission activities and health authority interactions, at a minimum the subgroups shown in table 2.3-4 are needed.

**Table 2.3-4 Regional/country subgroups to support regional/local submissions, for which selected tables have to be repeated**

| Preliminary* list of variable(s) / analyses to be repeated | Regional/country subgroup |
|---|---|
| **On FAS**<br>Disposition/Demography/Patient and disease characteristics<br>Incidence of primary endpoint<br>Incidence of key secondary endpoints<br>Incidence of secondary endpoints (includes death)<br>Analyses of other selected efficacy variables (e.g. hsCRP)<br>Medical environment (e.g. prior/concomitant medications including statin dose, revascularization/stenting/etc. for index MI)<br>[redacted]<br>**On SAF**<br>Adverse events by system organ class and preferred term<br>Serious adverse events by system organ class and preferred term<br>Deaths<br>Standardized MedDRA queries | 1. Japan<br>2. Japan and Korea<br>3. Mainland China<br>4. Mainland China + Taiwan<br>5. Asian patients: Japan, China, Taiwan, and Korea<br>6. India<br>7. Russia |

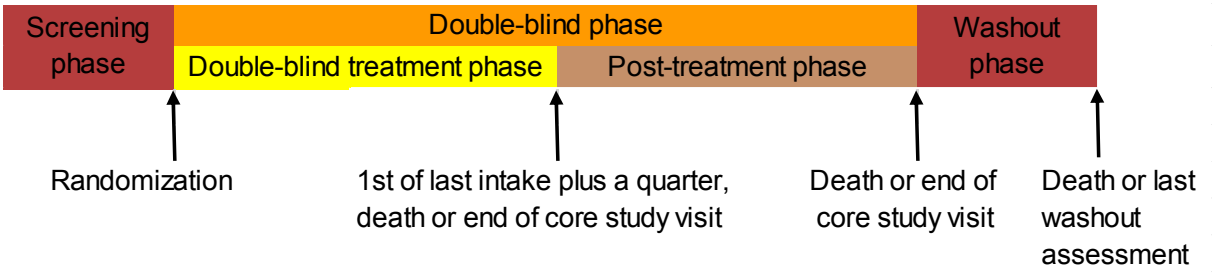
| Preliminary* list of variable(s) / analyses to be repeated | Regional/country subgroup |
|---|---|
| Adverse events of special interest | |

* subject to changes dependent on local requirements

## 2.3 Assessment windows, baseline and post baseline definitions, missing data handling

### 2.3.1 Assessment windows and visit mapping

**Figure 2-1 Assessment windows**




#### Screening phase

The screening phase is defined as the period prior and up to randomization. After obtaining written informed consent, the patient will be evaluated for eligibility to participate in to the study.

#### Double-blind phase

The double-blind phase includes double-blind treatment phase and post-treatment phase. It does not include the washout phase or any long-term follow-up after the completion of this trial. The double-blind phase defined here is the same as the double-blind treatment period defined in the study protocol.

#### Double-blind treatment phase

The double-blind treatment phase refers to the time period during which patients actually receive treatment (or are still expected to have significant PK exposure).

The double-blind treatment phase begins at the time of randomization and ends with the earliest of the last trial drug intake plus a quarter year (91 days), the death of the patient or the last study assessment in the main part of the trial (i.e. excluding the washout). During the double-blind treatment phase, patients will return for scheduled clinic visits.

#### Post-treatment phase

The post-treatment phase refers to the time period during which patients no longer receive treatment and no longer have significant PK exposure up to the end of the study visit for the

core study. As this is an outcome study, patients should nevertheless be followed-up during this phase until the end of the core study, in order to assess the occurrence of trial endpoints.

The post-treatment phase (after unscheduled drug discontinuation or study completion/termination) begins after last study drug intake plus a quarter year (91 days) + 1 day and ends with the date of the end of core study visit or the death of the patient. During the post-treatment phase, patients should also continue to return for scheduled clinic visits.

The distinction between double-blind treatment phase and post-treatment phase does not affect the primary intention-to-treat analysis, but affects e.g. the supportive on-treatment sensitivity analysis.

### 2.3.2 Baseline definition

Baseline is defined as the last available measurement during the screening phase, i.e. prior and up to randomization. That means that if the measurement at the baseline visit is missing, then the previous measurement preceding baseline will be used as the baseline measurement.

For hsCRP the baseline value will be the mean of the available hsCRP values within 60 days prior and up to randomization.

### 2.3.3 Post-baseline definition

Post-baseline is defined as any assessment after randomization including treatments and outcomes that occur after but on the same day as randomization. On the other hand, measurements and laboratory samples taken on the day of randomization during the randomization visit (at the end of which a patient is randomized) are considered baseline assessments.

### 2.3.4 On treatment definition

As baseline is relative to randomization, it is necessary to confirm where events/assessments occur relative to intake of dose. Any event/assessment occurring on or after the time of the first intake of IP and up to 119 days after the last treatment intake for those patients who discontinued treatment will be deemed to have occurred whilst “on treatment”.

**Scheduled Assessments:** For scheduled assessments, both date and time are expected to be recorded in the database. However in the instance that only the date is recorded for an assessment, then the assessment will be deemed to have occurred prior to treatment. This is justified as protocol indicates injections will be given after all other study assessments have been completed for each visit.

**Events:** Should an event occur on the same day as first intake of IP, then the most conservative approach is to consider this to be on treatment.

### 2.3.5 Visit mapping for the study completion visit

As the core study completion visit will occur after a different length of follow-up for different patients, it needs to be defined for which scheduled timepoint assessments at the study completion visit are counted. I.e. it needs to be defined when e.g. plotting the quarterly/bi-

yearly hsCRP values over time, whether the study completion hsCRP should count for the 0.25 year visit, the 0.5 year visit, the 0.75 year visit, the 1 year visit etc., which will be based on when it occurred.

When showing values over time, study completion visit assessments will be displayed as if they had occurred at the next scheduled timepoint for that assessment. For example, if a patient has a regular month 45 visit and then the study completion visit occurs 2 months later, then the study completion visit would be considered to provide the patient's values for month 48.

### 2.3.6 Data from unplanned visits and unscheduled visits

Data from unplanned visits and unscheduled visits will not be re-mapped to other visits for the purpose of displaying variables over time except for laboratory data. For laboratory data, a time window will be used to identify the assessment closest to the target day within the window over time. The unplanned and unscheduled visits will be used in the analysis of event endpoints (e.g. new onset diabetes), and will be counted towards shift tables or other analyses regarding the worst observed values for patients.

**Table 2-4 Time windows for endpoints based on HbA1c, FPG, standard hematology, biochemistry or fasting lipid profile**

| Time window | Start of time window (study days) | End of time window (study days) |
|---|---|---|
| Week 2 visit | 1, but only after the laboratory assessment of the baseline visit | 28 |
| Month 1.5 visit | 29 | 68 |
| Month 3 visit | 69 | 137 |
| Month 6 visit | 138 | 228 |
| Month 9 visit | 229 | 320 |
| Month 12 visit | 321 | 457 |
| Month X visit (where X≥18 is a multiple of 6) | Rounded [ ( 365.25 / 12 ) * ( X – 3 ) ] + 1 | Rounded [( 365.25 / 12 ) * ( X + 3 ) ] |

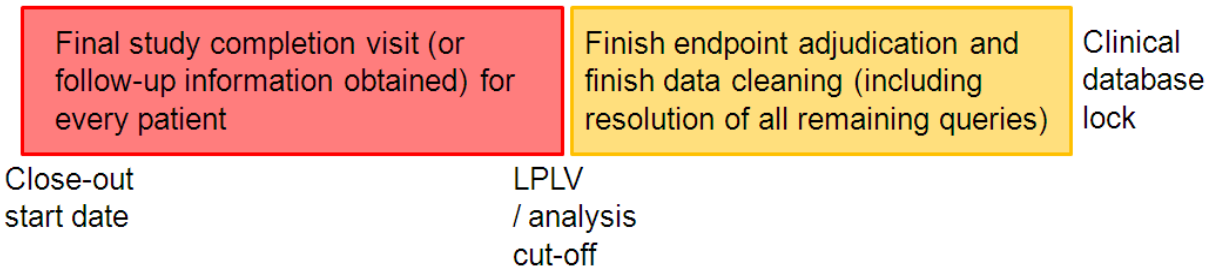
**Table 2-5 Time windows for urinalysis based endpoints**

| Time window | Start of time window (study days) | End of time window (study days) |
|---|---|---|
| Month 3 visit | 1, but only after the laboratory assessment of the baseline visit | 228 |
| Month 12 visit | 229 | 639 |
| Month 30 visit | 640 | 1096 |
| Month 42 visit | 1097 | 1461 |
| Month X visit (where X≥42 is a multiple of 12 added to 30) | Rounded [ ( 365.25 / 12 ) * ( X – 6 ) ] + 1 | Rounded [( 365.25 / 12 ) * ( X + 6 ) ] |

### 2.3.8 Close-out procedure

CANTOS is designed as an event driven trial with the aim to finish the core study when approximately 1,400 patients had a confirmed primary endpoint. A close-out procedure will be initiated with a timing appropriate to achieve this.

**Figure 2-2 Overview of close-out procedure**




The aim of the close-out procedure will be to allow for the execution of all necessary activities including the core study completion visits, data cleaning, raising and resolution of all queries to the clinical sites, and endpoint adjudication for endpoints that occurred during the core study and prior to final visit/analysis cut-off, whichever occurs first.

The close-out procedure will start based on the estimated date when the confirmed target number of events will be reached, with final patient visits occurring until the date of the last visit of the last patient (LPLV), and finish with the clinical database lock. The analysis cutoff date will be defined as the LPLV.

It will be attempted to follow each patient until a final study completion visit between the close-out start date and the analysis cut-off date unless the patient died. If a patient does not attend the final visit in person, there will be attempts to obtain information by other means between the close-out start date and the analysis cut-off date; this information will be available under the final study completion visit.

For patients who withdrew consent, the follow-up information such as vital status and whether endpoints occurred (“yes”, “no”, “unknown”) will still be collected during the close-out period to the maximum extent possible under local regulations.

Events occurring after the final visit, but before the analysis cut-off for patients that entered the long-term follow-up or the washout, will not be entered as endpoints into the database for the main trial.

### 2.3.9 Missing data handling

#### Imputation of missing date

As a general approach each time it is appropriate the partially missing date reported as character date in the data will be imputed.

- If only the month is known, then the 15th day of this month will be imputed.
- If only the year is known, then the 1st July will be imputed.

As a general approach no imputation will be performed on completely missing date.

For AEs specific rules to impute partially missing start/end dates are defined in the ADAM specification.

For prior and concomitant medications specific rules to impute partially missing dates are defined in the ADAM specification.

#### Time-to-event variables

Regarding time-to-event variables only observed/reported events will be used in the analysis – i.e. events will not be imputed for censored patients, except for specific sensitivity analyses. This means that in the primary analysis censoring of non-observed events will be assumed to be non-informative. Incomplete dates of events and censoring dates will be imputed as described below.

In the primary analysis, all patients, including those who discontinue study therapy due to lack of efficacy, adverse events or abnormal laboratory values, will be followed until death or the end of the study. Information on patients discontinuing study drug or participation in trial visits will be collected whenever possible and will be used in the analysis.

In general, every effort will be made to collect information about the primary outcome events for those patients who discontinued treatment or their participation in the trial. In patients who could not be followed up for primary outcome events, it is aimed to at least determine the vital status of the patients at the final visit.

The following rules will be applied separately for the composite MACE endpoint and for all its individual components. Patients who have not experienced the respective endpoint will be censored on the date of the last follow-up defined as the earliest of:

- the date of death unless the patient withdrew his consent for the collection of follow-up information,
- the date of the last visit satisfying at least one of the following:
  - yes to the question of "indicates if the subject attended the scheduled visit" on the Visit Information CRF
  - yes/no to the question of "any clinical events since last visit" on the Visit Information CRF

- yes/no to the top three questions (death, myocardial infarction, stroke) on the Clinical Event Tracking CRF.

Should the censoring date lie after the chosen analysis cut-off date, it will be set to the analysis cut-off date. Note that this censoring date should be derived independently of any specific analysis other than all cause mortality, because it will be used to describe the extent of follow-up for (potentially multiple) cardiovascular events and for multiple different time-to-event analyses (e.g. time to MACE, time to MI, analyses of multiple cardiovascular events). For all-cause death, the censoring will occur at the minimum of withdrawal of consent, last known alive date and analysis cut-off date.

If the date of a MACE endpoint or of censoring is not known or is incomplete following all attempts to get an approximate date, a day will be imputed using the following algorithm:

- If only the month of the event is known, then the 15th day of this month will be imputed.
- If only the year of the event is known, then the 1st July will be imputed.
- If year, month and day are unknown, the event date will be imputed as the randomization date and the patient will be assumed to have had a MACE event post-randomization on day 1 of the follow-up.
- If this imputation rule leads to a date before the randomization date or after a patient's last study visit or after a patient's death, but the date could have been on one of these dates, then the event date will be imputed to be the one out of these dates that the incomplete date is consistent with.
- In the first three bullet points it will also be taken into account, whether it is known that a patient was event free up to a certain time. If that is the case the imputed date will be restricted to be after the date up to which the patient was event free, unless this is incompatible with the incomplete date or violates the 4$^{th}$ bullet point.

### Date of trial drug injection

In case of missing information on the day of injection of any study medication the date of the corresponding dispensing visit will be used as date of injection. If that is missing the date foreseen for that visit in the protocol will be assumed.

### Other missing data

Missing laboratory, vital signs, ECG, quality of life and other data will not be imputed for the purpose of summary statistics. Such summary statistics will be displayed longitudinally showing only those values that are available. Subsequent more in-depth analyses may explore other approaches and appropriate data imputation methods will be chosen for each type of analysis.

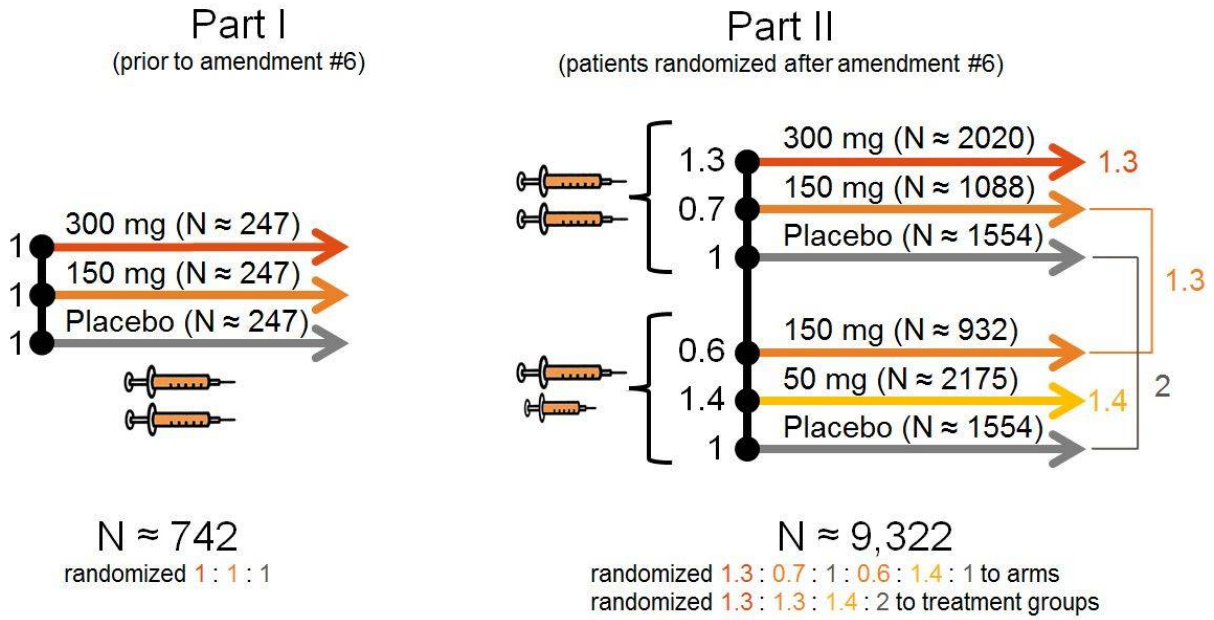
## 2.4 Key trial design aspects relevant to trial analysis

### Randomization and blinding

Initially, patients were randomized in a 1:1:1 fashion to 300 mg canakinumab with induction dose, 150 mg without an induction dose or placebo quarterly. Following Health Authority feedback on the doses a lower 50 mg dose arm was added in protocol amendment 6 in order to

examine the dose response of canakinumab in preventing recurrent cardiovascular events, and to determine if a dose lower than 150 mg would have a favorable risk benefit ratio. Following this amendment, randomized allocation to the four trial arms was to be unbalanced in order to optimize power, now that three active arms were compared versus a common control arm.

**Figure 2-3 Patient allocation to treatment arms prior and after protocol amendment 6 including reduced patient numbers following protocol amendment 8**




Protocol amendment 6 introduced 50 mg dose. In protocol amendment 8, the total sample size was changed from 17,200 to10,000. The allocation to active drug is always 2:1 (in Part I, Part IIA and IIB).

Prior to protocol amendment 6, the following two presentations of canakinumab solution for injection in pre-filled syringes are used in CANTOS:

- Canakinumab 150 mg/1 ml, and
- Placebo matching canakinumab 150 mg/1 ml

In addition, the following pre-filled syringes were available to be used for the purpose of implementing the 50 mg arm following amendment 6:

- Canakinumab 50 mg/0.5 ml, and
- Placebo matching canakinumab 50 mg/0.5 ml.

Canakinumab 50 mg/1 ml pre-filled syringes that fully match 150 mg pre-filled syringes were not available near the time of protocol amendment 6. Thus, in order to evaluate 300 mg, 150 mg and 50 mg versus placebo in a fully blinded fashion, a triple-dummy design with 3 injections per visit would have been required. As feedback from investigators indicated that this would constitute an unacceptable burden to patients we identified the approach shown in Figure 2-3 as the most appropriate solution for introducing the 50 mg dose into the CANTOS trial.

For the three trial arms in Randomization Plan A, study drug will consist of two injections of canakinumab 150 mg/1 ml or the matching placebo, while the three trial arms in Randomization Plan B study drug will consist of one injection of canakinumab 150 mg/1 ml, or the matching placebo and one injection of canakinumab 50 mg/0.5 ml, or the matching placebo. While investigators will be able to deduce from the size of the pre-filled syringes whether a patient is in Randomization Plan A or B, the chance that the patient is on active treatment is the same in either case (2:1). I.e. they are fully blinded with respect to active treatment or placebo, but can rule out one of the three possible active doses in case the patient was to be on active treatment.

As a result, the pre-specified primary analysis described in detail in section 6.3 comprises both

1. a randomized contrast between groups that are completely double-blind versus each other (canakinumab 150 mg versus placebo) and veiled versus the other groups and
2. two randomized contrasts between groups that are veiled (canakinumab 300 mg versus placebo and 50 mg versus placebo).

The term veiled is used in the sense of Senn (Senn 1995), who defines a veiled trial as a trial in which a patient's treatment group is not known, but the possibilities can be narrowed down to a subset of all existing treatment groups in the trial - in CANTOS every patient can clearly only be in 3 out of the 4 trial arms, either a patient can be on canakinumab 300 mg, 150 mg, or placebo or the patient can be on canakinumab 150 mg, 50 mg or placebo.

As far as combining data from trial part 1 and trial part 2 is concerned, the pre-specified analysis preserves the double-blind nature of the contrast in the same way that a patient-level (network) meta-analysis would. This is done by not making any direct comparisons between patients in trial part 2 and those in trial part 1 - i.e. no naïve pooling of patients across trial part, which would have been inappropriate, because patients were not concurrently recruited for the two trial parts and assignment to trial part was not at random. Instead, for example the 50 mg dose in trial part 2 can only be compared to the placebo group in trial part 1 indirectly using a double contrast by combining the 50 mg versus 300 mg and 50 mg versus 150 mg contrasts in trial part 2 with the 300 mg versus placebo and 150 mg versus placebo contrasts in trial part 1. As discussed by Senn, estimates obtained in this way have a substantially higher variance than the direct single (potentially biased) contrast (Senn 2008). Comparisons between treatments randomized to Part II on the other hand pool all concomitantly randomized arms because the potential bias is considered to be too small to justify the loss in efficiency incurred by such an approach. Furthermore, it was decided that any hypothetical bias could be handled for several reasons. Firstly, because patients are randomly assigned to the two randomization plans, one can expect patients in randomization plan A and B to be comparable at baseline. Secondly, any possible bias is expected to be small, if it exists. While investigators and patients can potentially identify whether patients are either on 300 mg, 150 mg, or placebo (trial part 1 or randomization plan A in trial part 2) or alternatively on 150 mg, 50 mg, or placebo (randomization plan B in trial part 2), the odds of a patient being on active treatment are the same in either case (2:1). In either case the number of potential treatment groups is only narrowed down to a limited extent, namely from 4 to 3. Taken together, this makes it unlikely that any influence of this knowledge on investigators and patients would have any substantial impact with respect to e.g. patient management, follow-up or endpoint

reporting. Thirdly, whether this is the case can be confirmed at the end of the trial due to the presence of both a placebo arm and a 150 mg arm in both randomization plans. For this purpose the placebo groups in randomization plans A and B will be compared against each other for the primary MACE endpoint, the 150 mg groups in randomization plan A and B will be compared and the 150 mg versus placebo contrasts will be compared between randomization plan A and B.

# 3 Patient disposition, background and demographic characteristics

## 3.1 Patient disposition

The number of patients screened, randomized and included in the full analysis set (FAS), the per protocol set 1 (PPS1), and the safety set will be presented by treatment group and overall for the screened set, as well as by trial part. For the screened set the reasons for screen failures will be tabulated. The number and percentage of patients in the randomized set who completed the study, who discontinued the study (lost to follow-up, patient/guardian decision, technical problems, death), the reason for discontinuation of study medication, and vital status (known vs unknown for lost to follow-up and patient/guardian decision) will be presented for each treatment group and all patients, as well as by trial part. Some patients may have multiple visits where drug is not administered prior to permanent discontinuation. Each visit may have a different reason for not administering drug. For these cases, the following hierarchy will be used to assign the most conservative reason for permanent discontinuation of study medication:

- If a patient has at least one AE with an outcome/ action reported as permanent treatment discontinuation, then AE will be the reason
- If the scenario in bullet 1 did not occur and there is at least one reason on the Dosage Administration Record CRF that states the reason for not administering study medication is “ADVERSE EVENT”, then AE will be the reason
- If the scenarios in bullets 1 and 2 did not occur, the reason associated with the last visit the patient actually came into the clinic will be used as the reason for permanent discontinuation of study medication if the patient did not receive study medication at that last in-clinic visit. If the patient did receive study drug at the last in-clinic visit, the reason for not administering study drug associated with the first visit that the patient did not come into the clinic will be used as the reason for permanent discontinuation of study medication.

The frequency (%) of patients with protocol deviations as well as the criteria leading to exclusions from analysis sets will be presented in separate tables for the randomized set. Finally, the number of enrolled and randomized patients by region (see section 2.2) as well as the number of patients enrolled and randomized per region and country will be presented descriptively for the randomized set overall and by trial part.

When the exact date of birth cannot be collected for data privacy reasons, the year of birth is collected instead and the patient’s age is imputed by the Oracle Clinical Remote Data Capture system (OC/RDC) on that basis. For that purpose OC/RDC assumes the patient’s birthday to

have been on 1 July of the year and calculates the patient's age as the number of days between that day and the date of the start of screening divided by 365.25.

## 3.2 Background and demographic characteristics

The common background and demographic variables will be summarized by treatment and the total of all patients using descriptive summary statistics (for continuous variables mean, median, standard deviation, Q1 ($25^{th}$ percentile), Q3 ($75^{th}$ percentile), minimum and maximum and for categorical variables frequency and percentage).

Summary of baseline demographic characteristics will include:

- Age [years]; age categories <65, ≥65 to <75, ≥75 years
- Sex (male/female)
- Race (Caucasian, Black, Asian, Native American, Pacific Islander, Unknown, Other) and Ethnicity (Hispanic, East Asian, Southeast Asian, South Asian, West Asian, Russian, Mixed ethnicity, Not reported, Unknown, Other)
- Height [cm]
- Weight [kg]
- Body mass index [$kg/m^2$] calculated as weight [kg]/ $height^2$ [$m^2$]; BMI categories of <25, ≥25 to <30, ≥30 to <35, ≥35 $kg/m^2$
- Waist circumference [cm]

Summary of baseline risk characteristics will include:

- Sitting pulse [bpm]
- Mean sitting SBP [mmHg]; additionally tertiles and categorization of < 130 vs. ≥ 130 mmHg
- Mean sitting DBP [mmHg]; additionally tertiles and categorization of < 80 vs. ≥ 80 mmHg
- Controlled baseline blood pressure (msSBP < 130 mmHg and msDBP < 80 mmHg)
- Smoking history (Never/Current/Former)
- Alcohol history (amount consumed on average per day <1, 1 to 2, >2 to 3, >3 to 4, >4 to 5, >5 or more)
- Cardiovascular risk factors and other co-morbidities (as per protocol solicited events CRF, unless otherwise specified):
    - Glycemic status (T2DM, T1DM, pre-diabetes, normoglycemic; MedDRA terms to identify T1DM will be pre-specified)
    - Complications of diabetes
        - Family history of diabetes
        - Diabetic retinopathy (Diabetic proliferative retinopathy/Diabetic non-proliferative retinopathy)
        - Diabetic nephropathy
        - Diabetic neuropathy
        - Diabetic foot ulcers

- Hypertension
- Dyslipidemia/Hyperlipidemia
- Prior TIA/stroke
  - Prior Transient Ischemic Attack
  - Prior Hemorrhagic Stroke
  - Prior Ischemic Stroke
  - Prior Stroke of unknown type
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- Peripheral Arterial Disease
- Prior repeated MI (multiple MIs identified via narrow Myocardial infarction SMQ in medical history)
- Prior coronary revascularization
  - Prior PCI (from the Prior Coronary Revascularization Procedures CRF)
    - For the qualifying MI
    - Not for the qualifying MI
  - [redacted]
  - Prior CABG (from the Prior Coronary Revascularization Procedures CRF)
    - For the qualifying MI
    - Not for the qualifying MI
- Heart failure (identified in Medical History using narrow Cardiac Failure SMQ)
- Medical history of gout (as per protocol solicited events CRF)
- Medical history of macular degeneration (list of MedDRA terms to be pre-specified)
- Time since index MI: summary statistics and classification into
  - < 6 months post-index MI sub-classified into
    - <30 days,
    - 30 days to 3 months (≤91 days)
    - 3 months (>91 days) to 6 months (<183 days)
  - ≥ 6 months post-index MI sub-classified into
    - 6 months (≥183 days) to 1 year (≤365 days)
    - 1 year (> 365 days) to 2 year (≤731 days)
    - 2 (>731 days) to 3 years (≤1096 days)
    - 3 (>1096 days) to 5 years (≤1826 days)
    - 5 (>1826 days) to 7 years (≤2557 days)
    - 7 (>2557 days) to 10 years (3653 days)

  - > 10 years (>3653 days)
- Type of qualifying MI (stemi, non-stemi, unknown)
- Type of diagnosis for qualifying MI (ECG, Elevated cardiac markers, Both, cardiac imaging)
- Number of prior MIs; number of prior MIs categorization as 0, 1, 2, 3, and ≥4
- hsCRP [mg/L]; tertiles, categories <3 mg/L, 3 to < 4 mg/L, ..., 9 to < 10 mg/L, ≥ 10 mg/L
- HbA1c [%]; categories <5.7%, ≥5.7 – <6.5%, ≥6.5 to <7.5%, ≥ 7.5% (overall and by each glycemic status)
- FPG [mg/dL]; categories <100 mg/dL, ≥100 to < 126 mg/dL, ≥126 to < 200 mg/dL, and ≥200 mg/dL (overall and by each glycemic status)
- Total cholesterol [mg/dL]; total cholesterol tertiles
- LDL-C mg/dL, LDL-C tertiles, LDL-C categories <70 mg/dL vs. ≥ 70 mg/dL, LDL-C < 100 mg/dL vs. ≥100 mg/dL.
- HDL-C mg/dL, HDL-C categories (<50 vs. ≥50 mg/dL for female; <40 vs. ≥40mg/dL for male)
- Triglycerides mg/dL; triglycerides tertiles; triglycerides categories < 150 vs. ≥ 150 mg/dL
- eGFR [mL/min/1.73m$^2$]; categories < 30, ≥30 to <60, ≥60 to <90 and ≥90 mL/min/1.73m$^2$
- Level of exercise (rarely/never, less than once a week, once a week, 2-3 times a week, 4-6 times a week, daily)
- Family history of MI, stroke, diabetes or premature coronary artery disease (No / Yes with sub-categorization by age at first event)
- Highest level of education (0 years, 1-4 years, 5-8 years, 9-12 years, 13-16 years, >16 years)

For purely descriptive purposes, treatment group comparability will be examined using the Cochran-Mantel-Haenszel test stratified by time since index MI (< vs. ≥ 6 months) and trial part for the categorical variables and the F-test with time since index MI (< vs. ≥ 6 months) and trial part as a categorical covariate for the continuous variables as appropriate. These p-values will be provided for descriptive purposes and will not be considered to define any formal basis for determining factors that should be included in statistical models. Any concerns about imbalances between treatment groups are already addressed by the pre-specified sensitivity analyses that assess the treatment effects adjusted for pre-specified potentially prognostic covariates (see Section 6.3).

[redacted]

## 3.3 Medical History

Any condition entered on the *relevant medical history / current medical conditions* CRF will be coded using the MedDRA dictionary. Frequency tables will be provided by primary system organ class, preferred term and treatment group.

Adverse events occurring between screening and baseline will also be reported as medical history, but additionally reported separately as adverse events occurring between screening and baseline.

# 4 Study medication

The duration of the double-blind treatment phase will be computed as the time from the first injection to the first out of

1. the last injection date plus a quarter year (91 days),
2. the patient's death
3. the patient's study completion visit during the study close-out period
4. or the analysis cut-off.

This algorithm reflects the planned treatment schedule and the long half-life of the study drug. Missing dates in this algorithm will be imputed in the same way as missing dates for time-to-event analyses.

The duration of the post-treatment phase will be computed as the number of days from the day after the end of the double-blind treatment phase to the end of the post-treatment phase. If the end of the double-blind treatment phase coincides with the end of the double-blind phase, then the duration of the post-treatment phase will be 0 days.

The duration of the double-blind phase and double-blind treatment phase will be summarized for the FAS and SAF by treatment group descriptively including by duration categories. Overall patient-years will also be displayed and will be computed as the sum of patient days for all patients divided by 365.25.The duration of treatment, and of exposure will be summarized for the SAF by treatment group descriptively including by duration categories. Overall patient-years of treatment will also be displayed and will be computed as the sum of patient days of treatment for all patients divided by 365.25.

Duration of exposure to study treatment excluding interruptions will be computed and summarized as above for the SAF, but not counting periods during which the last injection was more than a quarter year ago. Duration of exposure will also be summarized for patients who are alive at the end of the study.

An interruption is defined as the period of time in which IP is deemed to be inactive. In other words, if the date of the last dose + 91 days, falls prior to the following dosing date, then the difference between these is equivalent to the interruption duration. It should be noted that a single interruption may span multiple scheduled administrations of IP.

# 5 Concomitant medication

Prior or concomitant medications will be summarized for the safety set in separate tabulations based on the coding dictionary used. Medications will be presented in order of descending frequencies, by preferred terms and grouped by anatomical main group. Tables will show the overall number and percent of patients receiving at least one drug of a particular preferred term and at least one drug in a particular anatomical main group.

Prior medications and significant non-drug therapies are defined as any medications and significant non-drug therapies taken prior to the start of study drug. Concomitant medications and significant non-drug therapies are defined as those used during the double-blind phase. Concomitant medications that were prohibited as per protocol and given during the conduct of the study as well as significant non-drug therapies will be summarized.

Furthermore, classes of medications to be precisely defined in a separate Excel file that will be stored in the Novartis CREDI system, will be summarized separately. These will include classes such as the following:

- Anti-ischemic agents
  - Beta blockers
  - Cardiac glycosides
  - Intravenous or oral nitrates
  - Calcium channel blockers
- Antithrombotic agents (excluding antithrombotic intravenous and subcutaneous drugs intended for acute setting such as abciximab, eptifibatide, or tirofiban, which are not expected to be used chronically and are consequently less relevant in this population – these will be reported separately in the Clinical Study Report)
  - Acetylsalicylic acid (aspirin)
  - Non-aspirin oral anti-platelet agents
    - P2Y12 inhibitors (like clopidogrel, ticagrelor and prasugrel)
  - Direct Thrombin Inhibitors
  - Other antithrombotic agents (e.g. anticoagulants)
- Renin-Angiotensin-Aldosterone System inhibitors
  - Angiotensin Converting Enzyme (ACE) inhibitors (like ramipril)
  - Angiotensin II Receptor Blockers (ARBs)
  - Direct Renin Inhibitors
- Lipid-lowering agents
  - Statins
    - Atorvastatin
    - Rosuvastatin
    - Simvastatin
    - Other statins and statin combinations
  - Non-statins (like PCSK9 inhibitors, ezetimibe, fibrates, binding resins and nicotinic acid)
- Diuretics
  - Thiazide diuretics
  - Non-thiazide diuretics
- Anti-diabetic medications
  - Insulin

- Oral hypoglycemic agents
  - Thiazolidinediones
  - Metformin and other biguanides
  - Sulfonylureas
  - Alpha glucosidase inhibitors
    - sulfonamides
    - glymidine
  - DPP-4 inhibitors
  - Glinides
  - Other oral hypoglycemic agents (incl. SGLT2 inhibitors)
- GLP-1 analogs
- Aldosterone antagonists
- Proton pump inhibitors
- Treatments for macular degeneration

Additionally, the time course of the total daily dose level of concomitant statins will be reported at baseline and tracked over time. For each patient, the daily dose will be determined as follows:

1. For each patient, the daily dose will be determined by the statin use with start/stop date.
2. The daily dose by visit will be presented for the time course of the total daily dose level.
3. For each patient, the daily dose at each visit will be defined as the median of the daily doses since the last visit.
4. The median of individual patient's dose at each visit will be used as the point estimate for the plot/summary
5. In order to have consistent timings for the visit, the visits will be calculated based on protocol visit scheduling.

For this purpose doses of different statins will be equated according to Table 5-1. Median statin dose level, and patients on each standardized dose level will be summarized over time and displayed graphically, where possible. Patients will be classified according to whether they receive no, low dose, medium dose or high dose statins at baseline and during the study. This will be done separately by time since index MI (including finer categorizations than $<$ versus $\geq$ 6 months post index-MI) and by region. If new statins become available or substantial new information regarding their lipid lowering effects becomes available, this table may have to be updated.

**Table 5-1 Conversion table for statin doses**

| Dose intensity of statin | | Low | | | Medium | High | | |
|---|---|---|---|---|---|---|---|---|
| Standardized statin dose[1] [mg] (atorvastatin equivalent) | Scaling factor | 2.5 | 5 | 10 | 20 | 40 | 80 | 160[2] |

| Dose intensity of statin | | Low | | | Medium | High | | |
|---|---|---|---|---|---|---|---|---|
| Standardized statin dose[1] [mg] (atorvastatin equivalent) | Scaling factor | 2.5 | 5 | 10 | 20 | 40 | 80 | 160[2] |
| Atorvastatin (LIPITOR) | × 1 | | 5* | 10 | 20 | 40 | 80 | 160* |
| Lovastatin (MEVACOR) | / 4 | 10 | 20 | 40 | 80 | | | |
| Pravastatin (PRAVACHOL) | / 4 | 10 | 20 | 40 | 80 | | | |
| Simvastatin (ZOCOR) | / 2 | 5 | 10 | 20 | 40 | 80 | | |
| Fluvastatin (LESCOL) | / 8 | 20 | 40 | 80 | | | | |
| Rosuvastatin (CRESTOR) | × 2 | | 2.5* | 5 | 10 | 20 | 40 | 80* |
| Pitavastatin / itavastatin (LIVALO) | × 20 | | | | 1 | 2 | 4 | |
| Ezetimibe 10 mg + Simvastatin (VYTORIN, INEGY) | × 2 standardized statin dose | | | | 10/10 | 10/20 | 10/40 | 10/80 |
| Any other statin + ezetimibe or resin binders (e.g. colesevelam, cholestyramine, colestipol) | × 4 standardized statin dose | | | | | | | |

* off label dosing

[1] All other other drugs containing a statin should be reported for the statin dose only (including the following):
Statin + fenofibrate, + clopidogrel, + acetylsalicylic acid, + amlodipine, + ramipril

[2] Efficacy response curves flatten at the maximal equivalence of atorvastatin 160 mg or rosuvastatin 40 mg: further LDL lowering attempts with other lipid lowering drugs do not yield further LDL lowering. I.e. any calculated standardized statin dose > 160 mg will be considered as a standardized statin dose of 160 mg.

# 6 Efficacy evaluation

The primary analysis and all analyses of secondary/[redacted] variables will be conducted according to the intention-to-treat principle on the Full Analysis Set (FAS). Unless otherwise specified all time-to-event analyses will be based on events occurring during the double-blind phase. Unless otherwise specified "treatment groups" will refer to the four treatment groups canakinumab 300 mg, 150 mg, 50 mg or placebo. Comparing these separate active doses individually versus placebo will be the main method of analysis, but analyses of all canakinumab doses pooled versus placebo, as well as the active doses versus each other will also be conducted and dose response will be considered as described in section 6.5.

## 6.1 Variables

An overview of the primary, key secondary and other secondary variables as well as their components is given in Table 6.1. At the end of the study all endpoints will have been adjudicated by the Clinical Events Committee (CEC). The investigator reported events will only be considered as supportive. In order to assess the robustness of results at interim analyses the DMC will also be informed about events that have either been confirmed by adjudication, or reported by the investigator and not yet adjudicated. Events that have either been confirmed by the CEC or reported by the investigator but not yet been adjudicated will primarily be used for DMC activities to account for the time lag in adjudication.

**Table 6-1 Primary, key secondary and secondary variables for CANTOS**

| Endpoint (variable is time | Adjudication | Investigator | CEC confirmed or non- |
|---|---|---|---|

| to endpoint) | committee (CEC) confirmed | reported | adjudicated investigator reported ("unrefuted") |
|---|---|---|---|
| Major adverse cardiovascular events (MACE) | Primary | Supportive | Primarily for DMC monitoring |
| CV death | Supportive | Supportive | Primarily for DMC monitoring |
| fatal or non-fatal Myocardial infarction | Supportive | Supportive | Primarily for DMC monitoring |
| fatal or non-fatal stroke | Supportive | Supportive | Primarily for DMC monitoring |
| MACE or hospitalization for unstable angina requiring revascularization | Key secondary | Supportive | Primarily for DMC monitoring |
| hospitalization for unstable angina requiring revascularization | Supportive | Supportive | Primarily for DMC monitoring |
| New onset of diabetes | Key secondary | Primarily for DMC monitoring | Primarily for DMC monitoring |
| All-cause mortality | Secondary | Primarily for DMC monitoring | Primarily for DMC monitoring |
| All-cause mortality or myocardial infarction or stroke | Secondary | Supportive | Primarily for DMC monitoring |

### 6.1.1 Endpoint adjudication process

An external independent blinded adjudication committee will adjudicate all deaths, strokes, transient ischemic attacks, MIs, hospitalization for unstable angina requiring revascularization, new onsets of diabetes, all hospitalizations with adjudicatable information, and any other events that may be deemed appropriate for adjudication as the trial progresses.

As shown in Figure 6-1 there are 3 ways in which cases will be identified and forwarded to the adjudication committee.

The first identification process is by the investigator filling out an endpoint eCRF page. These pages trigger a notification to the Endpoint Coordinating Center ([redacted]) that a potential endpoint has occurred. [redacted] then assembles an adjudication package for the clinical endpoint adjudication committee (CEC).

The second identification process is by the sponsor clinical team reviews of the adverse events reporting via the clinical trial database, and ARGUS SAE reporting system.

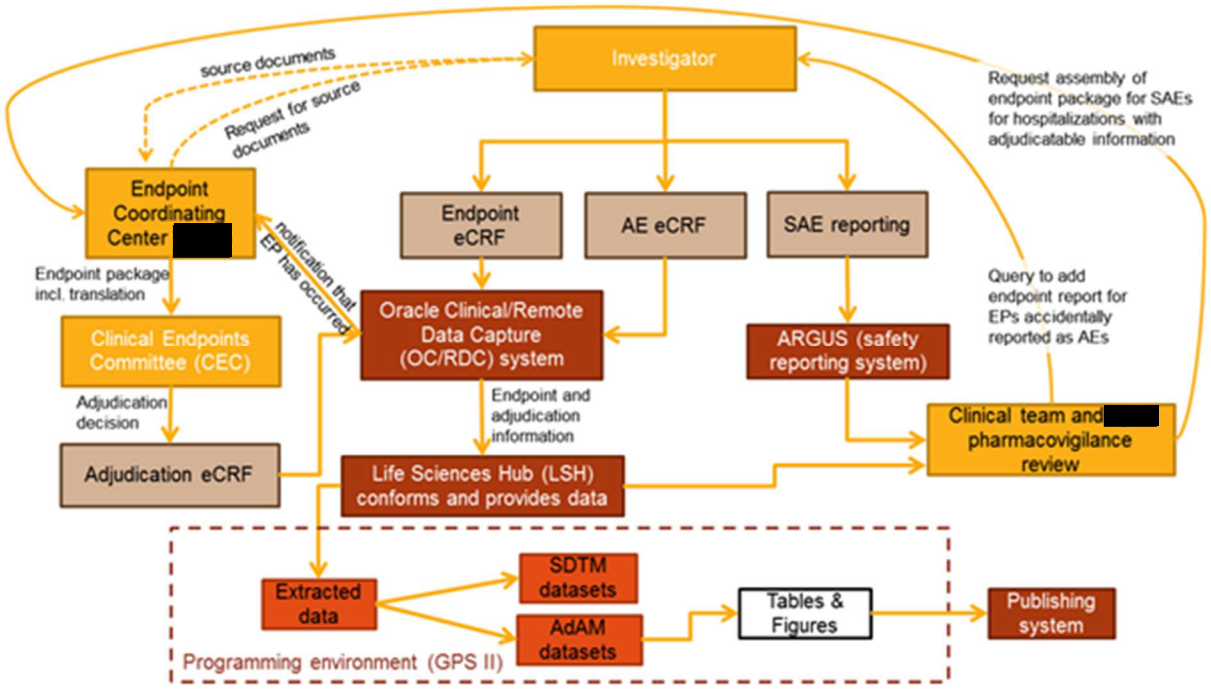
The third identification process is by [redacted] (a "third party") pharmacovigilence reviews of the Novartis ARGUS SAE reporting system of the SAEs which the sponsor clinical team did not send for adjudication.

This is done to ensure all hospitalizations with adjudicatable data are sent for potential MI or stroke adjudication. Adjudication packages for all SAEs requiring hospitalizations with adjudicatable material identified by these reviews that had not been previously been identified

by the investigator as a Study Endpoint, are also sent to the Endpoint Coordinating Center by Aptiv.

Once the Endpoint Coordinating Center ([redacted]) receives an endpoint notification, an adjudication packet consisting of source documents from the site is created and sent to the adjudication committee. The adjudication committee reviews the source documents, determines if the endpoint did or did not occur, and then fills out the adjudication eCRF page.

**Figure 6-1 Simplified overview of the adjudication process**




### 6.1.2 Primary efficacy variable

The primary efficacy variable is the time to first occurrence of a major adverse cardiovascular event (MACE) confirmed by Clinical Events Committee (CEC) adjudication, which is a composite endpoint consisting of CEC confirmed cardiovascular death, CEC confirmed MI, and CEC confirmed stroke.

The primary clinical endpoint MACE is a well-established endpoint for CV trials. An external Clinical Events Committee (CEC) will review and adjudicate all clinical events that constitute MACE and the key components of secondary endpoints on a blinded basis as described in section 6.1.1.

The individual components of MACE will be also analyzed as part of the primary efficacy analysis.

- Cardiovascular death
- Myocardial infarction (fatal or non-fatal)
- Stroke (fatal or non-fatal)

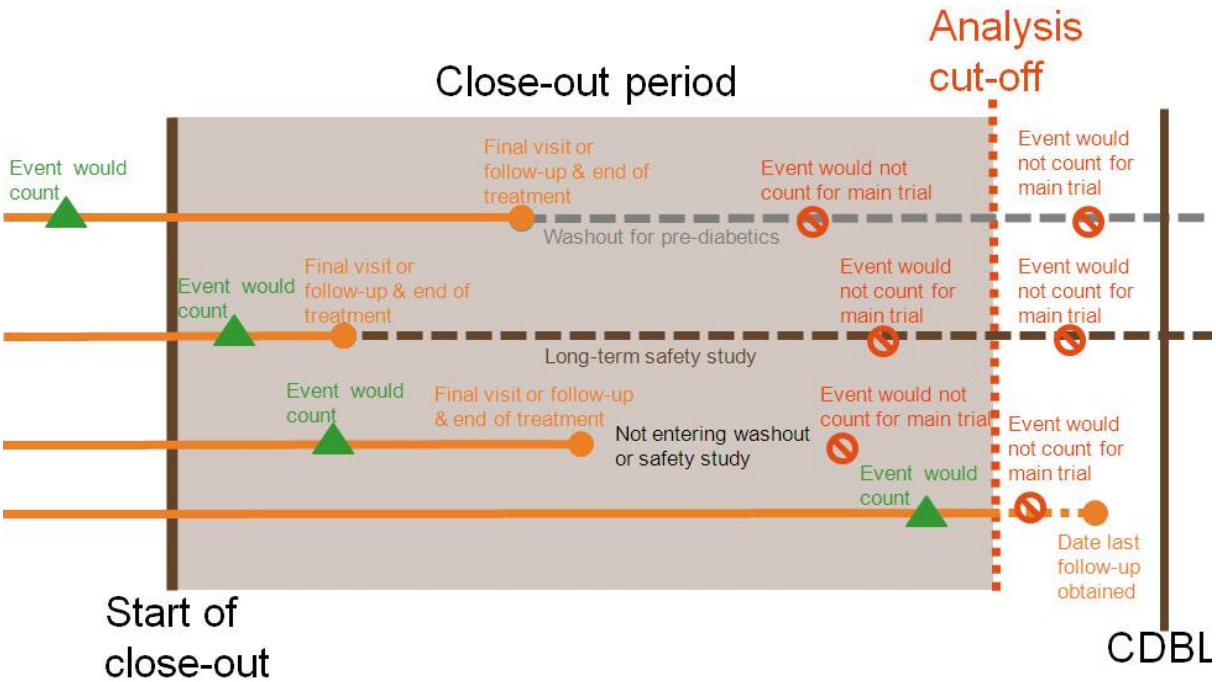
The time-to-event is computed as the number of days from randomization to the onset of the primary endpoint event. In the case of patients dying due to a MI, the date of the MI and not the date of death is considered to be the date of the endpoint. Data on patients who do not reach the primary endpoint by the study end date will be censored at the latest date they are known to be at risk.

Unless otherwise specified all time-to-event analyses will be based on events occurring during the double-blind phase (i.e. both the double-blind treatment phase and the post-treatment phase). This means that only events between a patient's randomization and the first of either

1. the patient's entry into the wash-out period (= end of study visit for the main study),
2. or the patient's entry into the long-term safety extension study (=end of study visit for the main study),
3. or for those patients proceeding to neither wash-out period nor safety extension the patient's end of study visit, but only if it occurred during the close-out period,
4. or otherwise the analysis cut-off

will be counted in these analyses as shown in Figure 6-2.

**Figure 6-2 Illustrative examples of events counting for the primary analysis**




Note: for patients entering the washout for pre-diabetics or the long-term safety study, events would be appropriately reported as part of the respective reports. Events occurring after study completion in patients not entering either washout or long-term safety study may still be subject to SAE reporting requirements and will be reported accordingly as per Novartis processes, if applicable.

### Reporting of type of first endpoint

The type of first endpoint will be identified as

- the first primary endpoint component (CV death incl. fatal MI or fatal stroke, non-fatal MI and non-fatal stroke) that occurs after randomization

- if more than one endpoint occurs on the same day, then for the purpose of reporting the type of the first endpoint a fatal event will be reported by preference (i.e. CV death in preference to a fatal-MI or fatal-stroke and these in turn in preference to a non-fatal MI or non-fatal stroke)
- should more than one endpoint fulfill the criteria above, then all of these will be reported (e.g. in case of both a non-fatal MI and non-fatal stroke on the same day "non-fatal MI and stroke" will be reported).

Note that the above only concerns the reporting of the type of the first endpoint. In the majority of tables the components of the primary endpoint will be shown separately and a patient may contribute to the numerator for no, one, two or all three components of the primary endpoint.

### 6.1.3 Key secondary efficacy variables

The following key secondary variables will be used in the analyses:

- Time to the first occurrence of an adjudication committee confirmed composite cardiovascular endpoint consisting of the components of the primary endpoint and hospitalization for unstable angina requiring unplanned revascularization
- Time to adjudication committee confirmed new onset of type 2 diabetes among those with pre-diabetes at randomization (i.e. excluding those that are normoglycemic at baseline)

### 6.1.4 Other secondary efficacy variables

The following other secondary efficacy variables will be used in analyses:

- Time to first event of MI, stroke, and all-cause mortality composite
- Time to all-cause mortality

Although all-cause mortality is considered a very important secondary endpoint due to its importance as both an efficacy and safety outcome, it is not part of the pre-specified testing procedure for primary and key secondary endpoints, as the expected number of deaths in the trial would not have ensured adequate power for this analysis.

## 6.2 Summarizing results across centers, countries and regions

As the target number of patients with a primary endpoint is not much larger than the number of centers, time-to-event analyses will be conducted on the pooled data from all centers without taking the center into account. However, the consistency of results by region will be explored.

## 6.3 Statistical hypothesis, model, and method of analysis

The primary analysis and all analyses of secondary/[redacted] variables unless otherwise stated will use the Full Analysis Set (FAS), which reflects the intention-to-treat principle.

The familywise type I error rate for this trial will be controlled at the one-sided 2.5% level. One-sided testing at the 2.5% level constitutes an equivalent level of evidence as two-sided testing at the 5% level. Two sided p-values will also be provided.

### 6.3.1 Primary variable

The primary statistical null hypotheses are

- $H_{11}$: The hazard rate of first adjudication committee confirmed MACE in the canakinumab 300 mg dose group is greater than or equal to the hazard rate of the placebo group
- $H_{21}$: The hazard rate of first adjudication committee confirmed MACE in the canakinumab 150 mg dose group is greater than or equal to the hazard rate of the placebo group
- H31: The hazard rate of first adjudication committee confirmed MACE in the canakinumab 50 mg dose group is greater than or equal to the hazard rate of the placebo group.

Each null hypothesis is tested against the one-sided alternative that the hazard rate is smaller for the respective active dose group than for the placebo group.

These hypotheses will be tested by comparing each dose to placebo with a log-rank test stratified by time since index MI (< 6 months and ≥ 6 months) and trial part on the full analysis set (FAS) according to the intent-to-treat principle. The stratification according to trial part entails some statistical inefficiency compared to a trial that would have had 4 treatment groups from the start, especially for the 50 mg dose, because comparisons against the placebo group in trial part 1 are only indirect via the comparisons versus other doses in trial part 2. However, this approach ensures that no biases arise e.g. due to differences over time between the two trial parts in the recruited patient population or background medical care. While investigators are not blinded to the assignment of patients to randomization plan A or B, this is not believed to introduce any biases with respect to the assessment of the primary endpoint and the standard or care provided to patients, because patients have an equal chance of being on an active treatment group under either randomization plan. Thus, it was not considered necessary to also stratify the analysis in this respect, but a sensitivity analysis will be conducted. A sensitivity analysis with respect to the consistency of treatment effects in the two parts of the trial will also be performed.

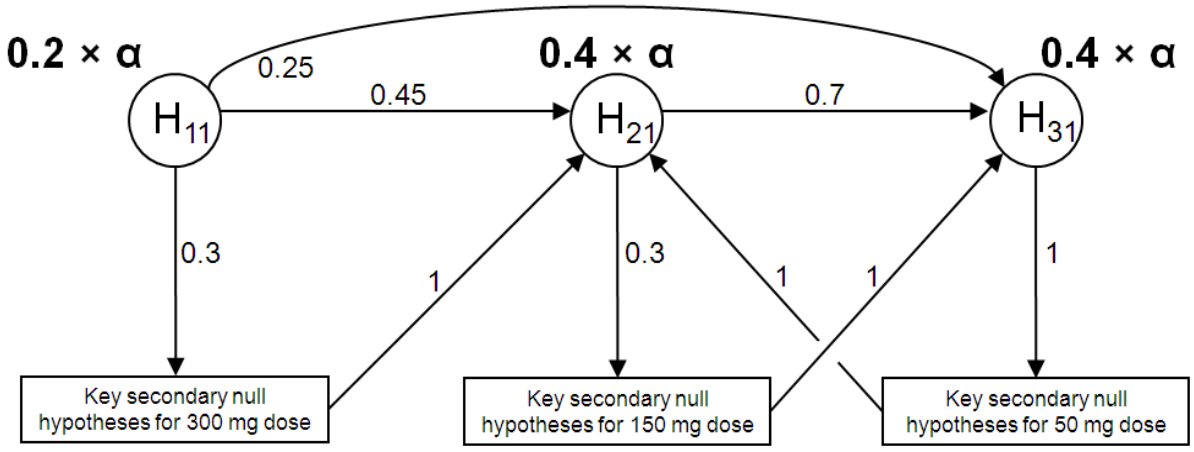
Note that no assumption of proportional hazards is necessary to test whether there is a higher probability of MACE in canakinumab than on placebo, because the employed stratified log-rank tests are valid even without proportional hazards. This assumption is only necessary to calculate estimated hazard ratios with confidence intervals for the effect of treatment (Bland and Altman 2004).

The family-wise error rate will be controlled at the two interim analyses and the final analysis using the closed testing procedure shown in Figure 6-4 based on the graphical method of Bretz et al. (Bretz, et al 2009); however, in intersection null hypotheses involving the primary null hypotheses for the 300 mg, 150 mg or 50 mg doses these primary null hypotheses will be tested using a weighted version of Dunnett's test (Dunnett 1955). Specifically this means that for any intersection hypothesis from the full closure that contains at least two of $H_{11}$, $H_{21}$ and $H_{31}$, a weighted Dunnett test amongst the primary null hypotheses is performed with the overall significance level for that test and the weighting chosen according to the weights assigned to these null hypotheses by the update algorithm of the graphical method. The nominal adjusted significance levels based on the weighted Dunnett test are always slightly larger than the corresponding Bonferroni levels would be. For example the mvtnorm package in R (Genz and Bretz 2009) calculates the nominal Dunnett significance levels at the final

analysis for the 300 mg, 150 mg and 50 mg doses versus placebo adjusting for the interim analysis in the test of the global null hypothesis to be 0.5500504%, 1.1001008% and 1.1001008%, respectively, as compared to Bonferroni levels of 0.49%, 0.98% and 0.98%. If other non-primary null hypotheses in such an intersection hypothesis have non-zero weight on the basis of Figure 6-4, the intersection null hypothesis will be rejected if either (a) these other null hypotheses can be rejected at a Bonferroni significance level based on that weight or (b) the primary null hypotheses can be rejected based on the weighted Dunnett test.

To illustrate the link between trial objectives and the close testing procedure, it may be helpful to consider what the procedure might have looked like, if the Dunnett test had not been used. This is shown in Figure 6-3. As discussed in detail below, introducing a Dunnett test required modifications with several additional vertices to ensure consonance (see Figure 6-4).

**Figure 6-3 Possible closed testing procedure if Dunnett test had not been used**




All other intersection hypotheses are tested with a weighted Bonferroni test. Protection of the family–wise error rate at level alpha is still guaranteed when the transition weights on the directed edges are chosen as in Figure 6-4 (see comment below the figure) and all the tests on secondary variables are performed at the level resulting from the graphical procedure. The consonance of the test procedure (Brannath and Bretz 2010) has also been ensured as described below.

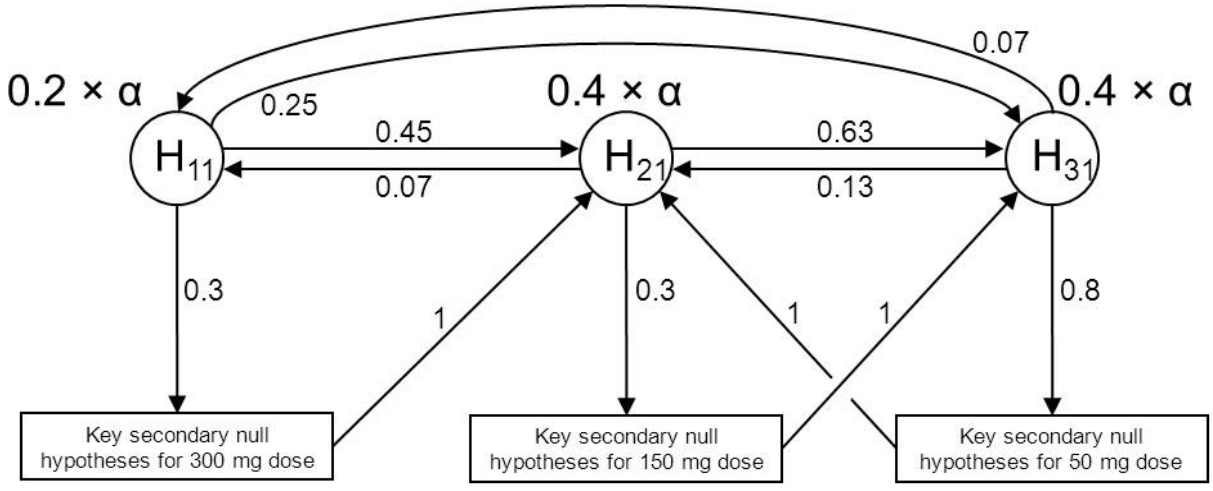
The testing procedure in Figure 6-4 initially splits the entire available significance level (at the final analysis 2.45%) between the three primary null hypotheses relating to the three doses (at the final analysis 20% of 2.45% = 0.49% for the primary null hypothesis of the 300 mg dose, 40% of 2.45% = 0.98% for that of the 150 mg dose and 40% of 2.45% = 0.98% for that of the 50 mg dose). These weights were chosen by balancing the prior expectations about the efficacy dose response relationship versus the potential for a better risk benefit ratio with lower doses. In particular, one aim was to ensure that there would be at least 80% power for the 50 mg dose to become significant assuming a 20% relative risk reduction for all doses. In the process of ensuring this, the weight for the 300 mg dose was reduced more than that of the 150 mg, because a higher dose is only of interest if it demonstrates a better efficacy than lower doses. This means that when the other doses are not effective the 300 mg dose would

only become significant with a relative risk reduction that is at least in the region of 17.5%, while for the 150 mg dose this would already be the case with relative risk reductions below 16%. However, this was considered acceptable in order to ensure the desired operating characteristics for the 150 mg and 50 mg dose.

Key secondary variables for a dose are tested using a weighted Bonferroni-Holm procedure (Holm 1979) only after successful rejection of the primary null hypothesis for that dose. In that case, a higher fraction of the local significance level that is passed from the primary null hypothesis for a dose to the key secondary endpoint for the same dose is assigned to the secondary CV composite endpoint (90%) than to the new onset of diabetes endpoint (10%). The details for the key secondary endpoints are more fully illustrated in Figure 6-5 in section 6.3.2, which discusses the key secondary efficacy variables.

Note that if the primary endpoint for the 300 mg dose (null hypothesis $H_{11}$) is rejected 45% of the local significance level assigned to that null hypothesis is shifted to the primary endpoint for the 150 mg dose (null hypothesis $H_{21}$) and 25% to the primary endpoint for the 50 mg dose (null hypothesis $H_{31}$). This reflects the possibility that lower doses could potentially have a better safety profile than higher doses; hence, it would be desirable to demonstrate the efficacy of lower doses even after demonstrating the efficacy of a higher dose. In contrast, the main reason why some of the local significance level assigned to lower doses is shifted to higher doses is to preserve the consonance of the test procedure; i.e. to avoid a situation in which e.g. the primary null hypothesis for the 300 mg would be rejected by the weighted Dunnett test, but could not be rejected by the chosen closed testing procedure due to insufficient alpha being assigned to some intersection null hypotheses.

**Figure 6-4 Closed testing procedure for primary and key secondary endpoints**




When the primary null hypothesis for a dose is rejected, some of the local significance level assigned to that primary null hypothesis is passed to the key secondary null hypotheses for the same dose. These are then tested using a weighted Bonferroni-Holm procedure at the available local significance level with weights of 90% for the key secondary CV composite and 10% for the key secondary new onset of diabetes endpoint. Only if both key secondary null hypotheses for a dose are rejected, will any of the local significance level assigned to these null hypotheses be passed on to null hypotheses for other doses. See Figure 6-5 for a version of the graph showing full details of the Bonferroni-Holm procedure for the key secondary endpoints.

Two efficacy interim analyses, at which the trial could be stopped for demonstrated efficacy, or one or more active arms could be stopped for futility, will be performed respectively after

50% and 75% of the target number of 1,400 patients have experienced a primary endpoint. Futility criteria and the criteria other than purely formal statistical significance required for stopping the trial for demonstrated efficacy will be specified in the Data Monitoring Committee charter. Protocol amendment 8 introduced an additional early futility analysis to be performed at approximately 30% of the target number of 1400 patients have experienced CEC confirmed MACE. The analysis plan of the early futility was discussed in a separate document.

A fixed Bonferroni split of the one sided significance level will be used to account for the two efficacy interim analyses and the final analysis, with a significance level of 0.01% for the first and 0.04% for the second efficacy interim analysis. I.e. the closed testing procedure will be performed with an one-sided significance level of 0.01% at the first efficacy interim analysis, with an one-sided significance level of 0.04% at the second efficacy interim analysis and with an one-sided significance level of 2.45% at the final analysis. In this fashion the familywise type I error rate will be controlled at the overall (one-sided) significance level $\alpha = 2.5\%$.

No multiplicity adjustment will be made for the additional early futility analysis, because it will not result in rejecting a null hypothesis for efficacy superiority.

The appropriate nominal weighted Dunnett significance levels will be calculated e.g. using the mvtnorm R package (Genz and Bretz 2009). For the example of the global null hypothesis this can be done using R code (or corresponding SAS code) similar to the following

```
library(mvtnorm)
alpha <- 0.0245 # level at which Dunnett test is performed
rho <- 0.5 # assumed correlation of test statistics
corr <- (1-rho)*diag(3)+rho # correlation matrix
weight <- c(0.2,0.4,0.4) # weights for hypotheses in Dunnett test
alphaw <- alpha*weight
MVN <- function(alpha1, w, corr, alpha){
      x<-qnorm(1-alpha1*w)
      return(1-pmvnorm(upper=x,corr=corr) - alpha)
      }
opt3 <- uniroot(MVN, lower = alpha, upper = 2*alpha, w=weight,
corr=corr, alpha=alpha, tol=1E-12)
critwDun <- qnorm(1-opt3$root*weight)
nominal <- opt3$root*weight
```

The hazard ratios and their associated confidence intervals will be estimated by means of a (Cox 1972) proportional-hazards model stratified by time since index MI (< 6 months, ≥ 6 months) and by trial part (trial part 1, trial part 2) using treatment (canakinumab doses and placebo) as a factor in the model. This will be done using SAS code similar to the following (or equivalent code in another programming language such as R):

```
proc phreg data=dataset;
  class treatment(ref="Placebo") <time since index MI> <trial part>;
```

```
    model time*event(0) = treatment / ties=exact risklimits;
    strata <time since MI > <trial part>;
run;
```

Should one of the three active arms be stopped due to safety reasons or futility, then thereafter the pre-specified testing procedure will be performed for the other arms treating all null hypotheses for the stopped arm as non-rejected. For patients whose treatment is stopped as a result of a data monitoring committee recommendation to suspend treatment in one dose group, follow-up for cardiovascular (CV) and safety events will continue. Since these patients will not receive further study medication, less frequent follow-up may be conducted every 6 months. Continued follow-up would likely enhance understanding of the safety of the other remaining dose groups as well. Follow-up will simply continue in all scenarios until 1,400 patients have had a primary endpoint and the only difference to the pre-planned study conduct will be that the patients in the stopped treatment arms are not being treated after the DMC decision point.

### Supportive, sensitivity and subgroup analyses for the primary endpoint

The components of the composite primary efficacy endpoint (CV death, fatal or non-fatal MI, fatal or non-fatal stroke) will also be analyzed individually in order to evaluate their contributions to the overall treatment effect. No adjustment for multiplicity is foreseen for this purpose.

The three pooled canakinumab doses will be compared to placebo on the primary composite endpoint and its components, because in case of similar efficacy on all three doses this would be the most powerful test of the scientific hypothesis addressed by this trial. Additionally, when all three doses have similar efficacy, such a pooled analysis approach may provide the most precise information on the effects in subgroups.

Dose response will be evaluated using the methods discussed in section 6.5. Additionally, comparisons between the canakinumab doses based on the primary model will also be reported.

Kaplan-Meier type analyses will be presented as described in section 6.4 overall and by trial part to summarize the time to first event in the composite endpoint, by presenting the time-dependent cumulative frequency and percentage of patients who reach the primary composite endpoint by treatment group.

Additionally, log time by treatment interaction will be evaluated to identify violations in the proportional hazards assumption, which might indicate a time-varying treatment effect. An Anderson-Darling type test based on a score process (cumulative sum of the Schoenfeld residuals) as recommended by Kvaløy and Neef (Kvaløy and Neef 2004) will be performed. For each trial arm within each trial part –log(S(time)) vs. time and log(–log(S(time)) ) vs. log(time) will be plotted.

To ensure that any unblinding of patients during the double-blind phase had no meaningful impact on the overall study results, a sensitivity analysis of the primary endpoint will be performed by excluding data from all patients who were unblinded during the double blind phase of the study (i.e. all the unblinded patients will be censored at the time of unblinding).

The primary variable and its components will also be analyzed on the PPS1. Additionally, in an on-treatment analysis on the FAS, patients will be considered censored at the latest 28 days after the end of the double-blind treatment phase, which corresponds to up to 119 days after the last study drug injection based on the long elimination half-life of the drug. 91 days constitute a bit over 3 half-lives and 28 days add an additional half-life. In the on-treatment analysis patients will stay in the risk set until this point unless their censoring date occurred earlier and will not leave and re-enter the risk set due to treatment interruptions. Besides adjudicated endpoints investigator reported outcomes will also be analyzed on the FAS.

Once the results become available a covariate adjusted analysis may be considered. This covariate-adjusted Cox-regression model will consider the following pre-specified baseline covariates: linear age, sex, race, linear BMI, smoking status, region as defined in section 2.2, linear baseline log hsCRP, type of qualifying MI (STEMI/NSTEMI/Unknown), glycemic status, duration of diabetes for diabetic patients, hypertension, linear baseline systolic blood pressure, linear and quadratic baseline diastolic blood pressure, history of heart failure, dyslipidemia, linear baseline LDL-C, linear baseline HDL-C, linear baseline triglycerides, linear number of prior MIs, prior PCI, prior TIA/stroke, linear eGFR, no/low or medium dose/high dose statin use at baseline as defined in section 5, baseline aspirin therapy, baseline P2Y12 therapy, RAAS inhibitor use, beta blocker use, to account for other factors that might affect prognosis.

Furthermore, the pre-specified subgroup analyses will be conducted for the subgroups listed in section 2.2. Results will be presented graphically as forest plots. The objective of the subgroup analyses is to evaluate the consistency of treatment effects across a wide variety of patient groups. For this purpose the availability of data from three different doses should be helpful, because most subgroup effects would be affect all doses to some extent, which provides an additional way of identifying spurious findings that occurred by chance. Additional subgroup analyses will be considered and potentially pre-specified prior to unblinding of trial database for final analysis. It is clear that due to the number of subgroups a substantial number of totally spurious numerical imbalances in treatment effects between subgroups has to be expected – including some with an interaction p-value $\leq 0.05$ and/or with apparently no or reversed effects. The association between changes in hsCRP and outcomes will also be explored.

To assess whether pooling randomization plan A and B in trial part 2 is appropriate, the placebo groups in randomization plans A and B will be compared against each other, the 150 mg groups in randomization plan A and B will be compared and the 150 mg versus placebo contrasts will be compared qualitatively between randomization plan A and B.

As a further sensitivity analysis in this respect, randomization plan will be used as an additional stratification factor in the primary model. This analysis will use SAS code similar to the following (where randomization plan would be a 3 level stratification factor with 1 level for trial part 1 and two further levels for each randomization plan in trial part 2) or equivalent code in another programming language such as R:

```
proc phreg data=dataset;
  class treatment(ref="Placebo") <time since index MI>
      <randomization plan>;
```

```
  model time*event(0) = treatment treatment*<randomization plan>
      / ties=exact risklimits;
  strata <time since index MI> <randomization plan>;
  hazardratio treatment / diff=ref cl=both;
run;
```

A sensitivity analysis with respect to the consistency of treatment effects in the two parts of the trial will also be performed. This analysis will use SAS code similar to the following or equivalent code in another programming language such as R:

```
proc phreg data=dataset;
  class treatment(ref="Placebo") <time since index MI> <trial part>;
  model time*event(0) = treatment treatment*<trial part>
      / ties=exact risklimits;
  strata <time since index MI> <trial part>;
  hazardratio treatment / diff=ref cl=both;
run;
```

### Calculation of multiplicity adjusted one- and two-sided p-values

Multiplicity adjusted p-values will be calculated for each null hypothesis as the smallest significance level at which one can reject that null hypothesis using the testing procedure shown in Figure 6-4 while also taking into account the interim analyses. For this purpose the algorithm described by Bretz et al. (2009) will be applied to the unadjusted one-sided p-values and the resultant p-values (adjusted only for the multiple endpoints and doses) will be adjusted for the interim analyses by scaling them by a factor of 250 (=0.025/0.0001) at the first efficacy interim analysis, 62.5 (=0.025/0.0004) at the second efficacy interim analysis and 0.025/0.0245 (≈1.02) at the final analysis. Because a weighted Dunnett test will be used for any intersection hypothesis from the full closure that contains at least two of $H_{11}$, $H_{21}$ or $H_{31}$, the algorithm of Bretz et al. (2009) needs to be extended. For this purpose the weights $w_{11}$, $w_{21}$ and $w_{31}$ used as part of that algorithm would be defined as the respective Dunnett level divided by the overall significance level at the respective analysis.

For example, for the global null hypothesis instead of $w_{11}$=0.2, $w_{21}$=0.4 and $w_{31}$=0.4 as for a Bonferroni test based procedure – for which $w_{11} + w_{21} + w_{31} \leq 1$ would always hold – instead $w_{11} + w_{21} + w_{31} > 1$ would hold in this case. For example, using the nominal Dunnett significance levels for 300 mg, 150 mg and 50 mg doses at the final analysis calculated using the mvtnorm package in R as described earlier in section 6.3.1, one obtains $w_{11}$ = 0.5500504% / 2.45% ≈ 0.2245103, $w_{21}$ = 1.1001008% / 2.45% ≈ 0.4490207 and $w_{31}$ = 1.1001008% / 2.45% ≈ 0.4490207.

In case health authorities, journals or the data monitoring committee require two-sided p-values, these will be provided in a manner that firstly ensures that the test decision on the rejection of the primary and key secondary null hypotheses remains unchanged and that secondly uses in principle the same testing procedure for evaluating the potential superiority of placebo over each dose of the test drug.

On this basis, the same test procedure as shown in Figure 6-3 will also be conducted in a one-sided fashion on the 2.5% significance level to test for the superiority of placebo over each canakinumab dose arm. The multiplicity adjusted two-sided p-values for each canakinumab dose arm will then be calculated as the minimum out of

- 2 × the multiplicity adjusted one-sided p-value for the superiority of that canakinumab dose arm over placebo,
- 2 × the corresponding multiplicity adjusted one-sided p-value for the superiority of placebo over that canakinumab dose arm versus
- and 1.

**Exploration of dose response relationship and the effects of induction dosing in terms of cardiovascular endpoints**

The method described in section 6.5 will be applied to assess dose response on cardiovascular endpoints over the whole duration of the trial.

It will be explored whether the early high induction dose regimen in 300 mg canakinumab arm has an impact on early (within 90 days of randomization) primary and secondary clinical cardiovascular events when compared to the placebo arm and to the 150 mg and 50 mg canakinumab arms, which have no early high induction dose regimen. For this purpose the primary endpoint and its components will be assessed over 90 days post-randomization. Similarly, the primary endpoint and its components will also be assessed at 6 months and 1 year, repeatedly.

### 6.3.2 Key secondary efficacy variables

The following hypotheses will be tested with respect to the key secondary variables for the canakinumab 300 mg dose versus placebo

- $H_{12}$: The hazard rate of first adjudication committee confirmed secondary composite CV endpoint in the canakinumab 300 mg dose group is greater than or equal to the hazard rate of the placebo group
- $H_{13}$: The hazard rate of adjudication committee confirmed new onset of diabetes for pre-diabetic patients in the canakinumab 300 mg dose group is greater than or equal to the hazard rate of the placebo group

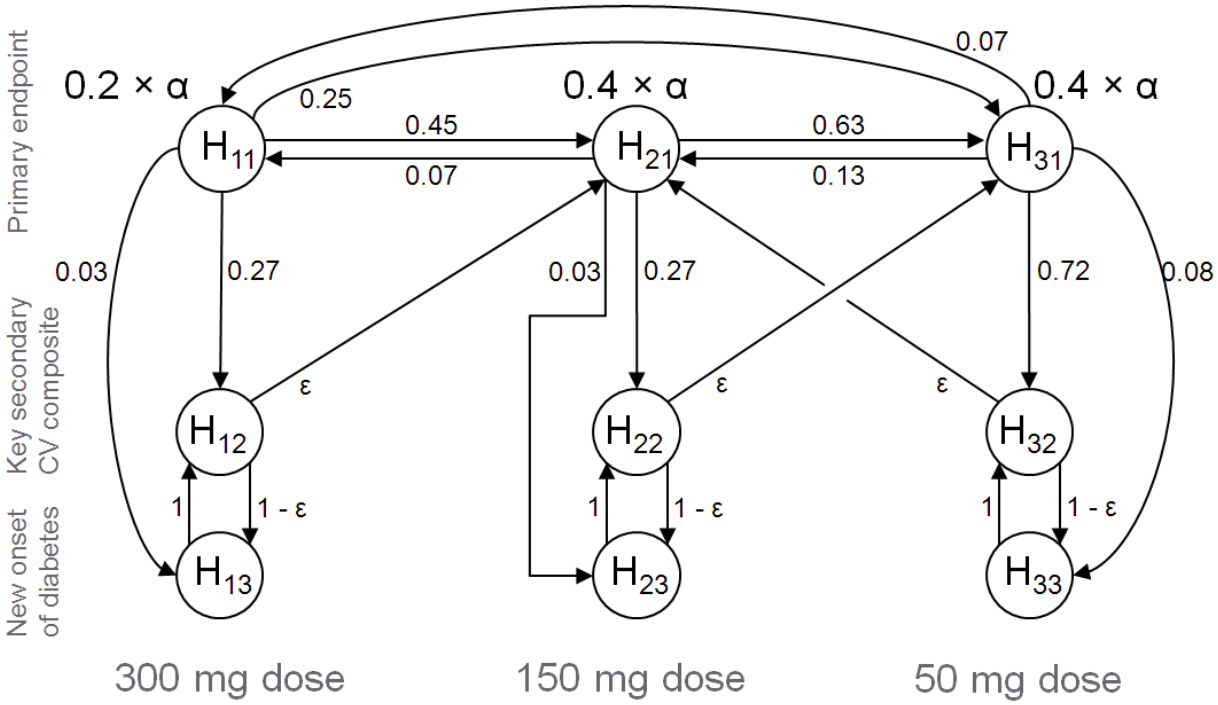
Each null hypothesis is tested against the one-sided alternative that the hazard rate is smaller for the canakinumab 300 mg dose group than in the placebo group. The corresponding hypotheses for the comparison of the canakinumab 150 mg dose versus placebo are $H_{22}$ for the secondary composite CV endpoint and $H_{23}$ for the new onset of diabetes endpoint. For the 50 mg dose they are $H_{32}$ for the secondary composite CV endpoint and $H_{33}$ for the new onset of diabetes endpoint.

All key secondary efficacy variables will be analyzed on the FAS and PPS1with a log-rank test stratified by time since index MI and trial part. The hazard ratios will be estimated using a Cox regression model stratified by time since index MI and trial part.

Kaplan-Meier plots showing each treatment will be provided overall and by trial part as described in section 6.4. The multiplicity adjustment used to protect the familywise type I

error rate is shown in Figure 6-4 and showing the full details of testing the key secondary endpoints in Figure 6-5. For each dose analyses of key secondary endpoints will be conducted even if the primary null hypothesis for the dose has not been rejected, but formally no rejection of the null hypotheses for the key secondary endpoints will be possible in that case.

**Figure 6-5 Closed testing procedure with full details of testing of key secondary endpoints**




This figure as compared to Figure 6-4 shows the full details of the Bonferroni-Holm procedure for the key secondary endpoints in a graphical form, but is otherwise identical in content.

Based on this testing procedure, once the primary null hypothesis for a dose has been rejected the key secondary endpoints for that dose are tested using a weighted Bonferroni-Holm test (Holm 1979) at the available local significance level for the key secondary endpoints for that dose. The weighting of this Bonferroni-Holm procedure will be 90% for the key secondary CV composite and 10% for the key secondary new onset of diabetes endpoint.

The secondary efficacy variable corresponding to new onset diabetes in patients with pre-diabetes at randomization will be the time from randomization to the first of repeated FPG $\geq$ 126 mg/dL or the first of repeated HbA1c $\geq$ 6.5% or start of new anti-diabetic concomitant medication(s) for glucose lowering purpose. All identified cases of new onset of diabetes will be confirmed by the adjudication committee.

Due to the discrete nature of the visits when new onset of type 2 diabetes can be determined, events identified at the same visit time point as defined in section 2.3.6 for different patients will be considered as tied events and the exact method for handling ties will be used. This assumes that for each of these patients new onset of diabetes occurred at some time point since the previous visit, but that due to the impossibility of continuous monitoring of the patients the true order in which each of them progressed to diabetes is unknown (Allison 1995,

pp. 127 to 137). For this variable patients without new onset of diabetes will be considered censored at the time of their last laboratory assessment. Patients not diagnosed at a regular visit or at the study completion visit will be assigned to one of the scheduled visits as per the algorithm in section 2.3.6.

Subgroup analyses as specified in Table 2-3-1 will be conducted for the key secondary new onset of diabetes endpoint. These will include subgroup analyses by race, which could potentially influence HbA1c levels.

Given some patients may have medical conditions that elevate their HbA1c levels that are unrelated to diabetes (e.g. patients with hemoglobinopathies such as sickle cell disease, while several other conditions causing interference such as e.g. uremia, liver disease, alcoholism and drug (opiate) use are excluded from this protocol), a second set of new onset diabetes criteria will be defined. This second set of criteria will determine new onset diabetes for these patients based on having either (2 consecutive HbA1c >= 6.5 % and 2 consecutive FPG values >= 126 mg/dL from the same visits as HbA1c) or (HbA1c >= 6.5 % at one visit reported and FPG >= 126 mg/dL reported at same visit and FPG>=126 confirmed at the consecutive visit) or 2 consecutive FPG values >= 126 mg/dL. The identification of these patients will be done in a blinded fashion by a medical monitor using medical history data and other baseline data. All patients deemed not to have any of these medical conditions will be assessed for new onset diabetes as described in the first set of criteria (see paragraph above). New onset of diabetes data will be reported overall and separately for these patients in whom diagnoses based on HbA1c levels is not appropriate (defined as “FPG >= 126 mg/dl at 2 occasions” or “Confirmed prescription and use of Diabetes Drug” based on the New Onset Diabetes (Adjudication) CRF page) The analysis described above will also be performed using this alternative new onset diabetes definition.

### Supportive and sensitivity analyses for key secondary endpoints

To support the analysis for the new onset of diabetes endpoint, the criteria defining new onset will be presented, the diagnostic values causing diagnosis summarized and sensitivity analyses conducted based on each criterion alone.

The three pooled canakinumab doses will be compared to placebo on the key secondary endpoints. Additionally, comparisons between the canakinumab doses will also be reported using the pre-specified log-rank test with hazard ratios reported based on Cox regression.

Additionally, Kaplan-Meier plots will be provided separately for each of the followings:

- Time to adjudication committee confirmed new onset of type 2 diabetes among those with normoglycemic at randomization
- Time to adjudication committee confirmed new onset of type 2 diabetes among those with either pre-diabetes or normoglycemic at randomization
- Time to first pre-diabetes among those with normoglycemic at randomization.

The key secondary cardiovascular endpoint will also be analyzed adjusting for the same covariates used for the covariate adjusted analysis of the primary variable.

The key secondary endpoint new onset of diabetes will also be supported by analyses using data from the study’s washout period. These analyses will be defined in a separate analysis

plan for the washout period and will be reported in a separate study report or addendum to the main study report.

### 6.3.3 Other secondary efficacy variables

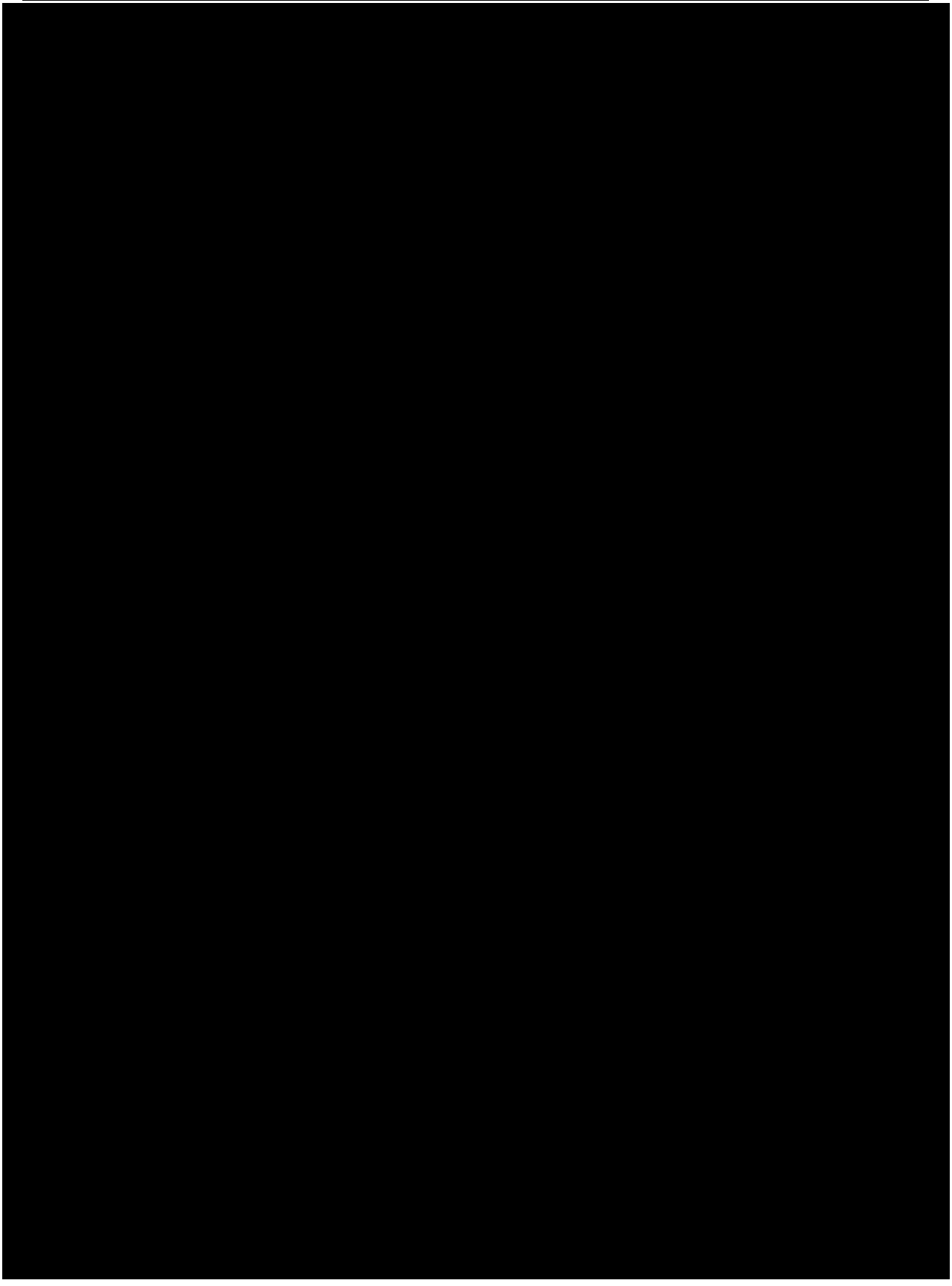
As stated in section 6.1.3 all-cause mortality is considered a very important secondary endpoint due to its importance as both an efficacy and safety outcome, but it is not part of the pre-specified testing procedure for primary and key secondary endpoints, because given the expected number of deaths in the trial it would not have been possible to adequately power the key secondary mortality endpoint.

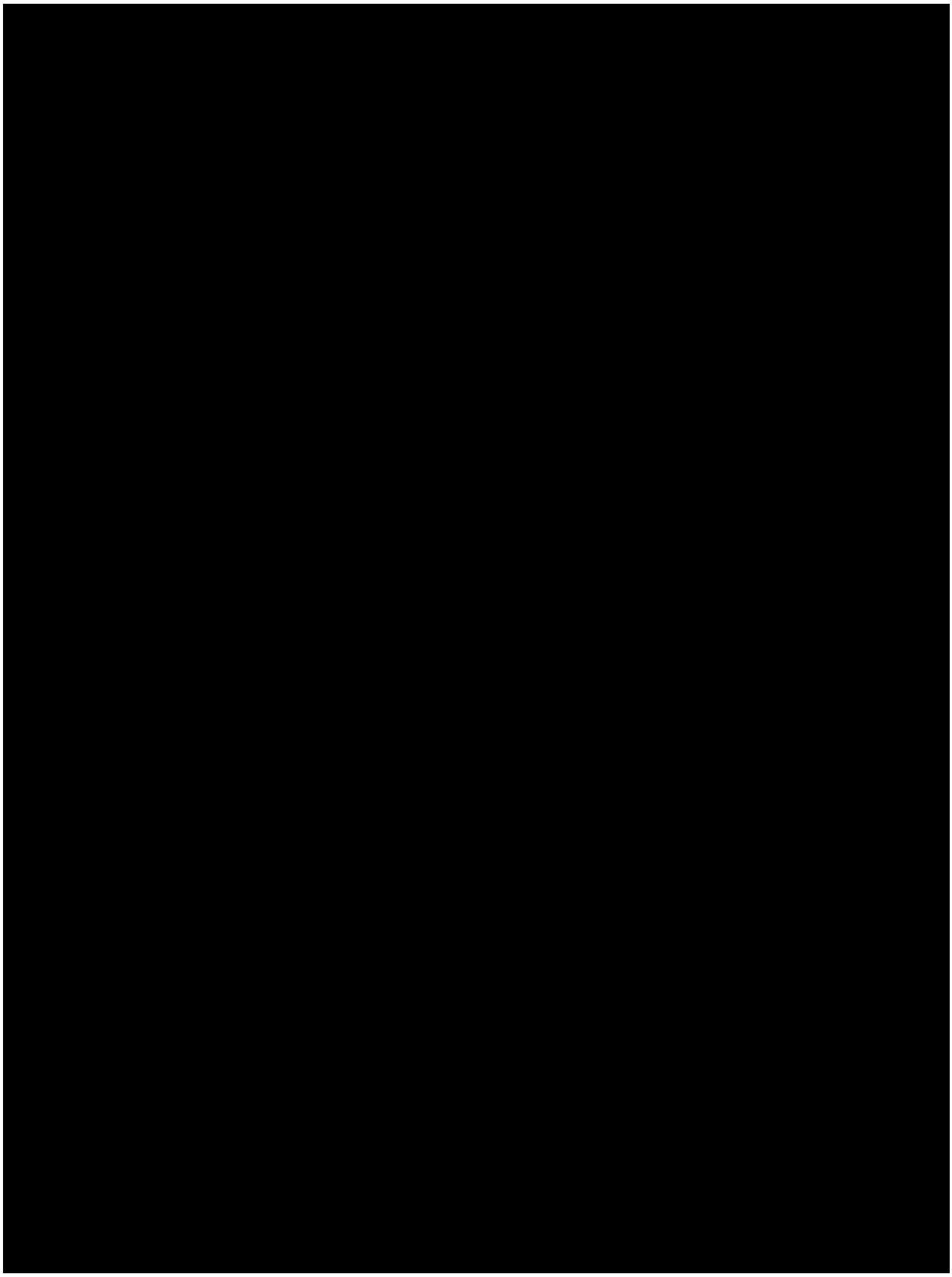
All-cause death and the composite of all-cause death, stroke or MI will be analyzed on the FAS and PPS1with a log-rank test stratified by time since index MI and trial part. The hazard ratios will be estimated using a Cox regression model stratified by time since index MI and trial part.

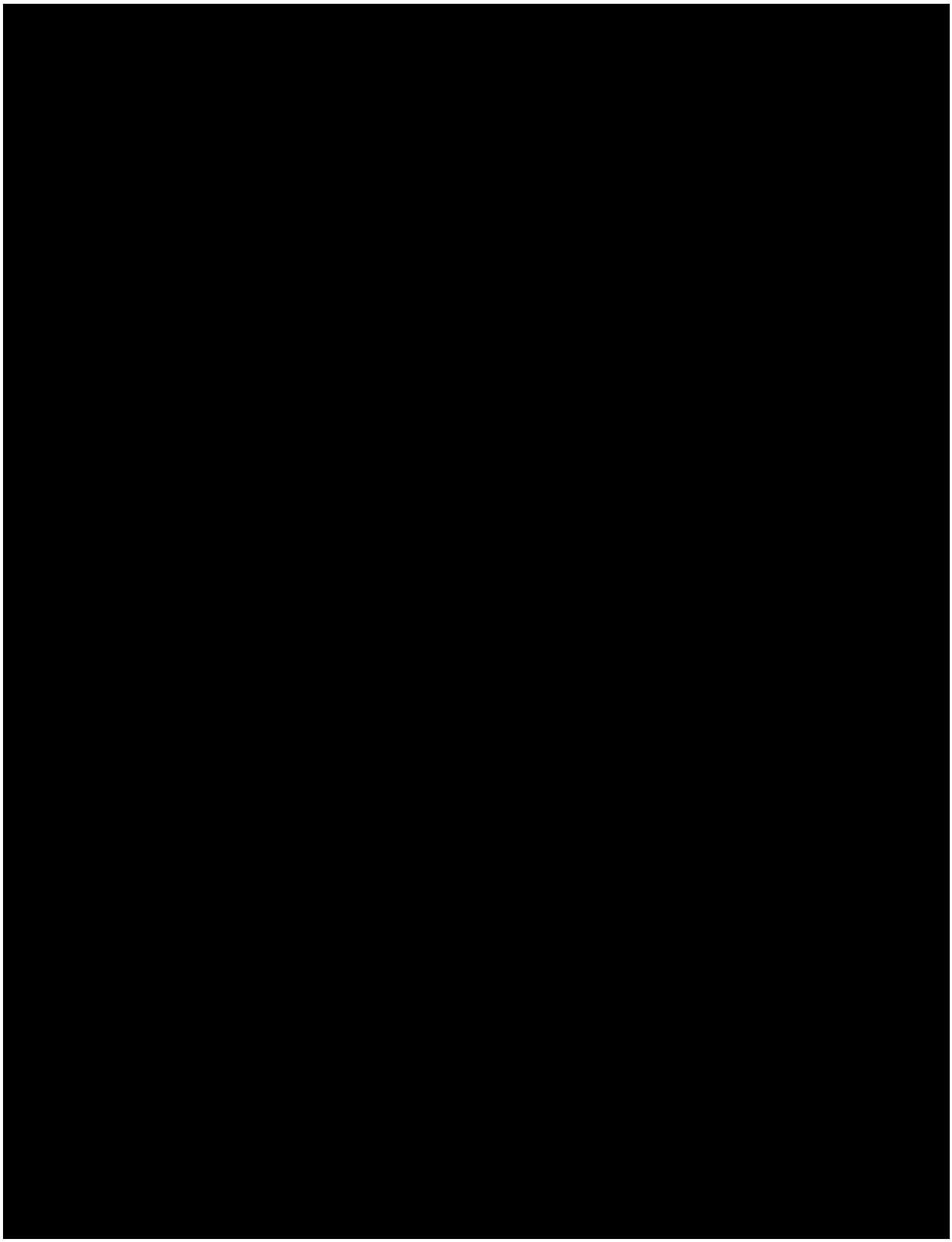
Kaplan-Meier plots showing each treatment will be provided both overall and by trial part. Patients who did not die will be considered censored at the last time they were reported to be alive.

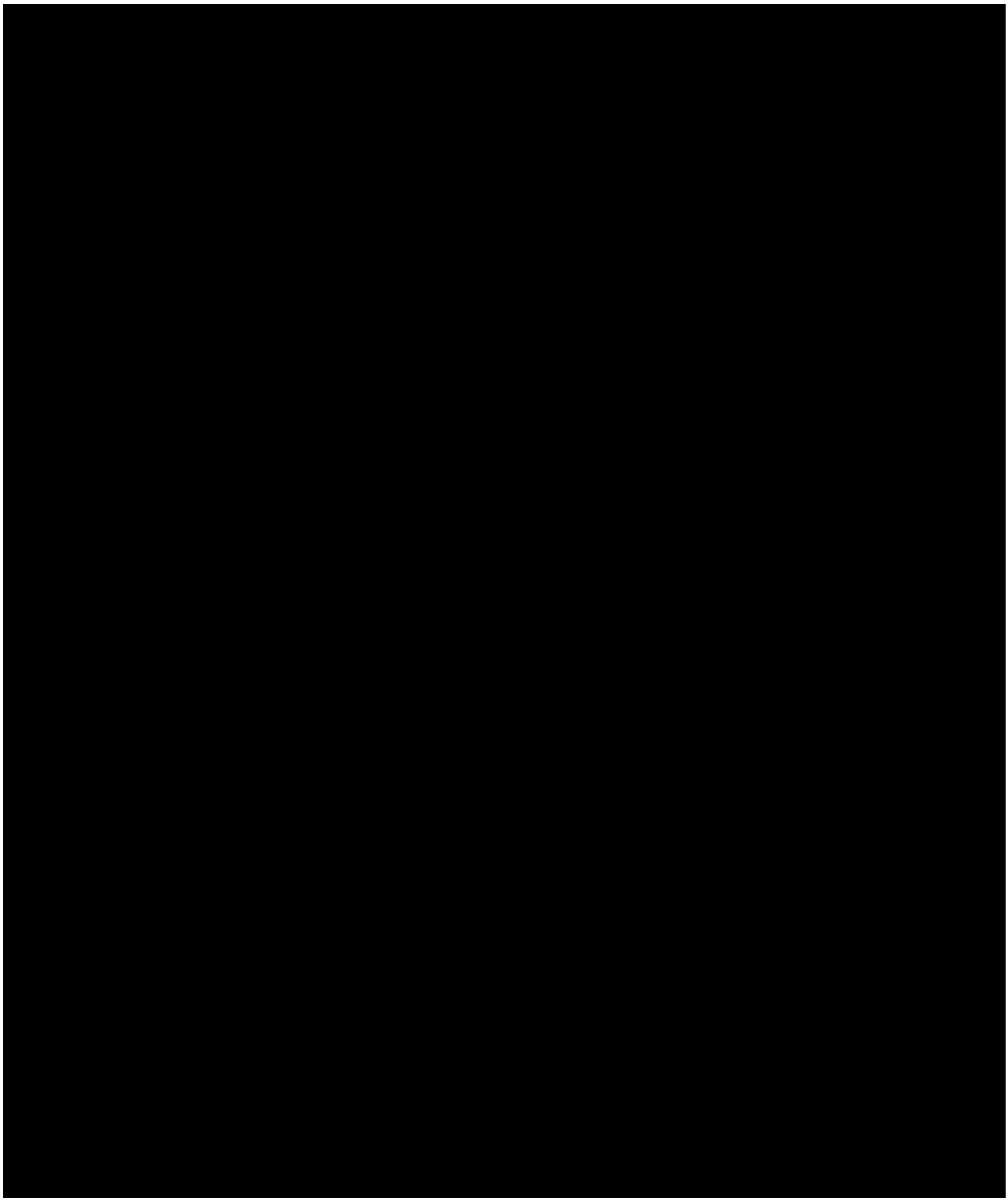
These secondary endpoints will also be analyzed adjusting for the same covariates used for the covariate adjusted analysis of the primary variable.

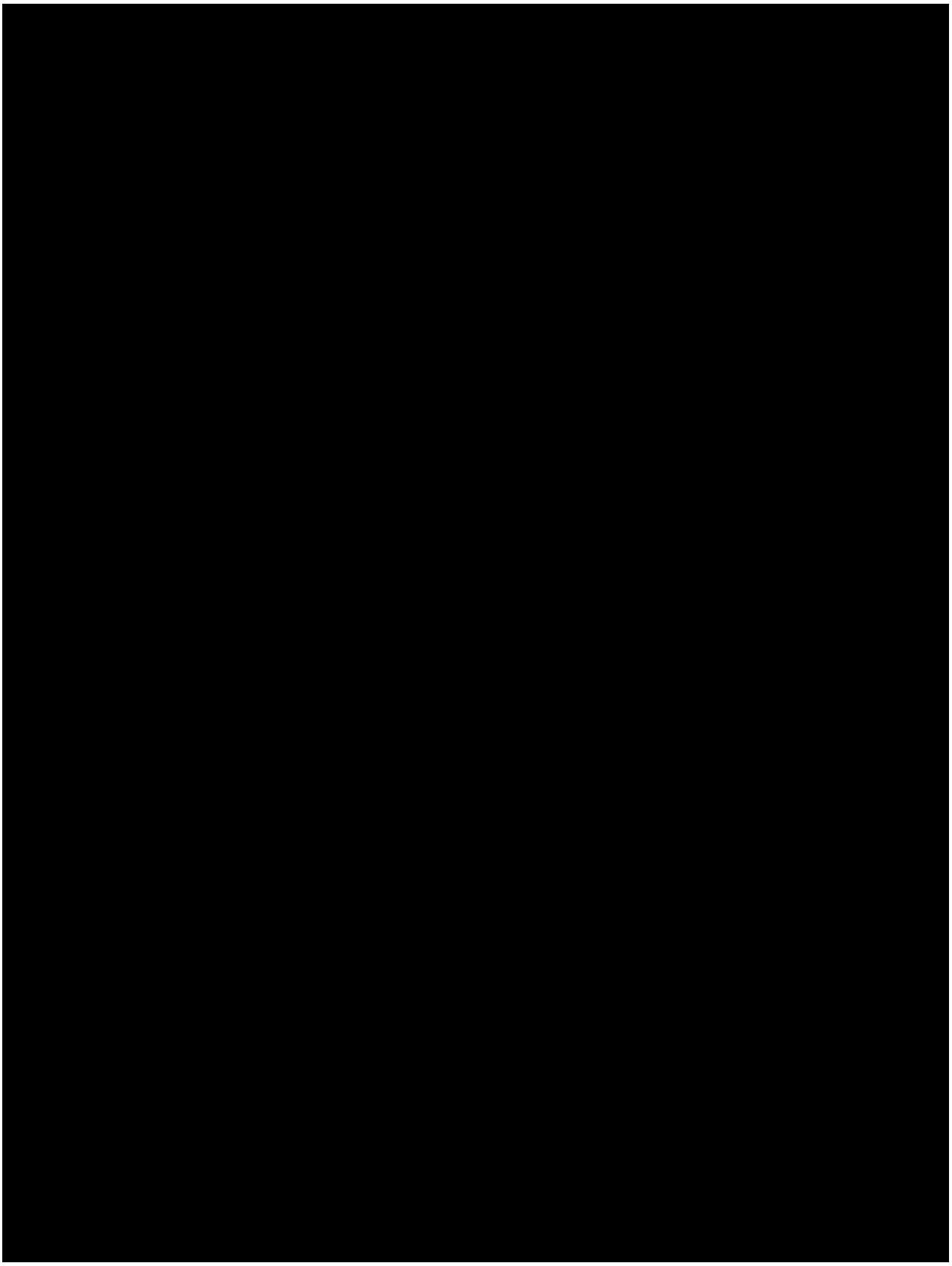
The three pooled canakinumab doses will be compared to placebo on the other secondary endpoints. Additionally, comparisons between the canakinumab doses will also be reported.

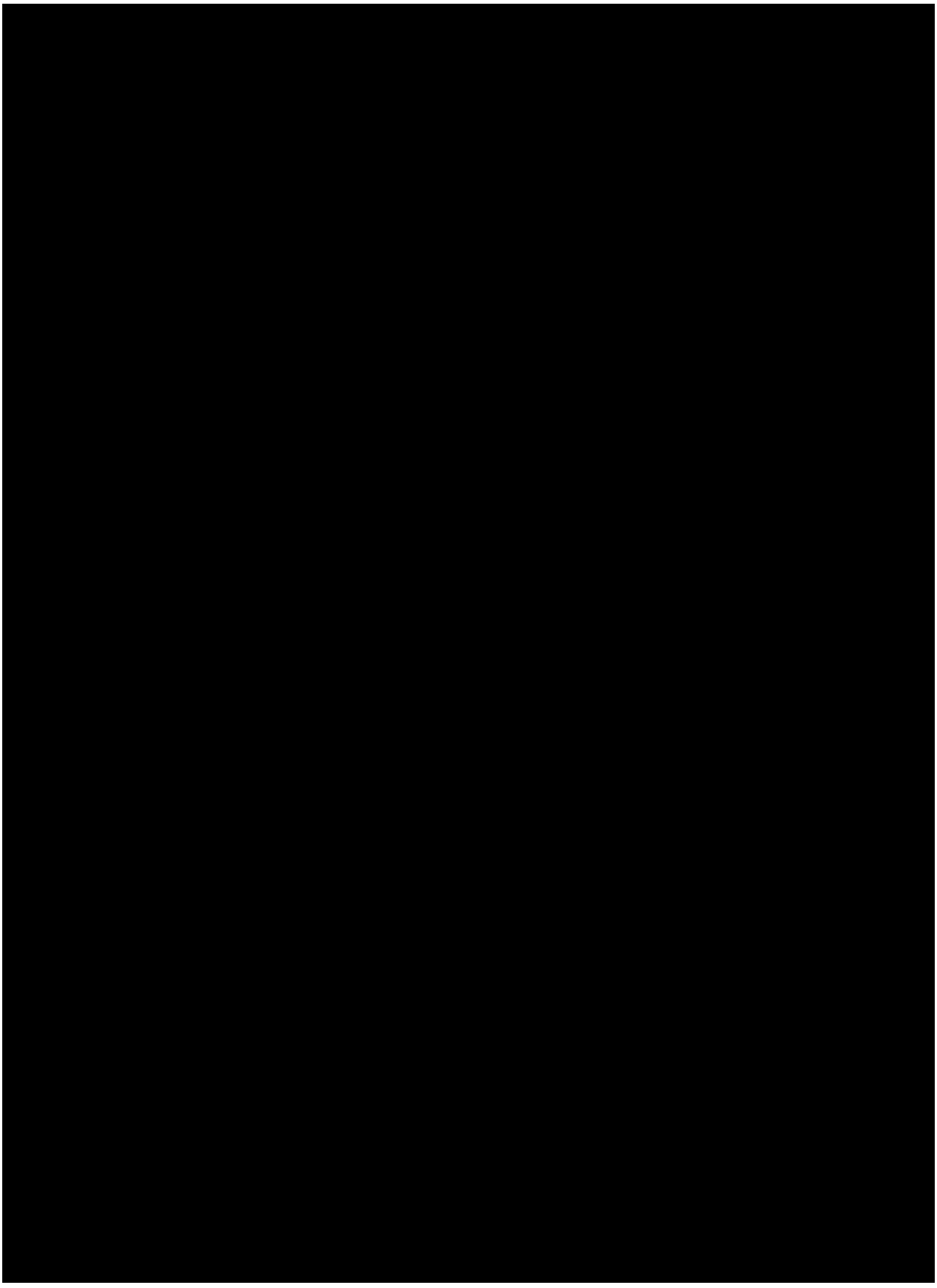


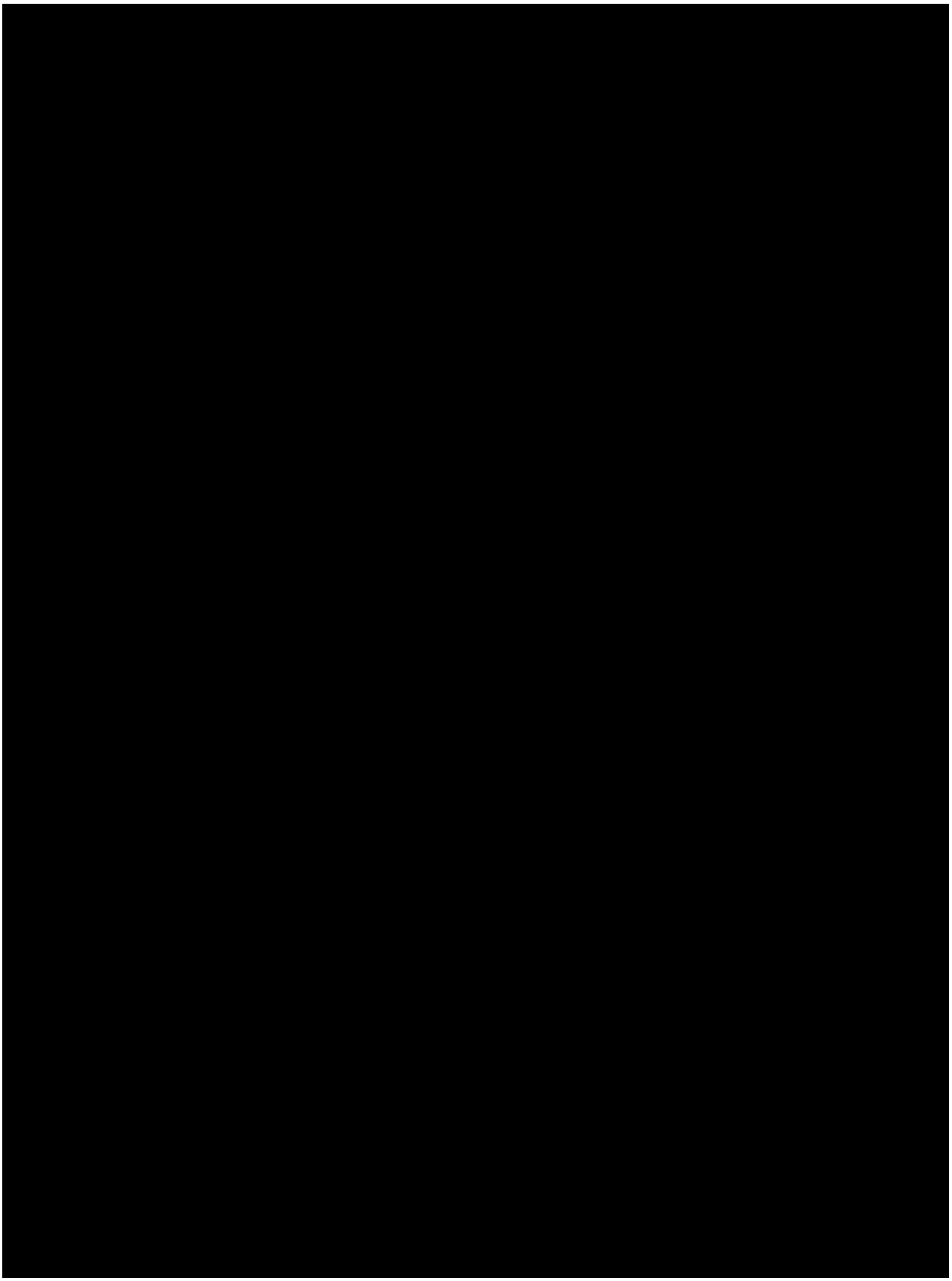


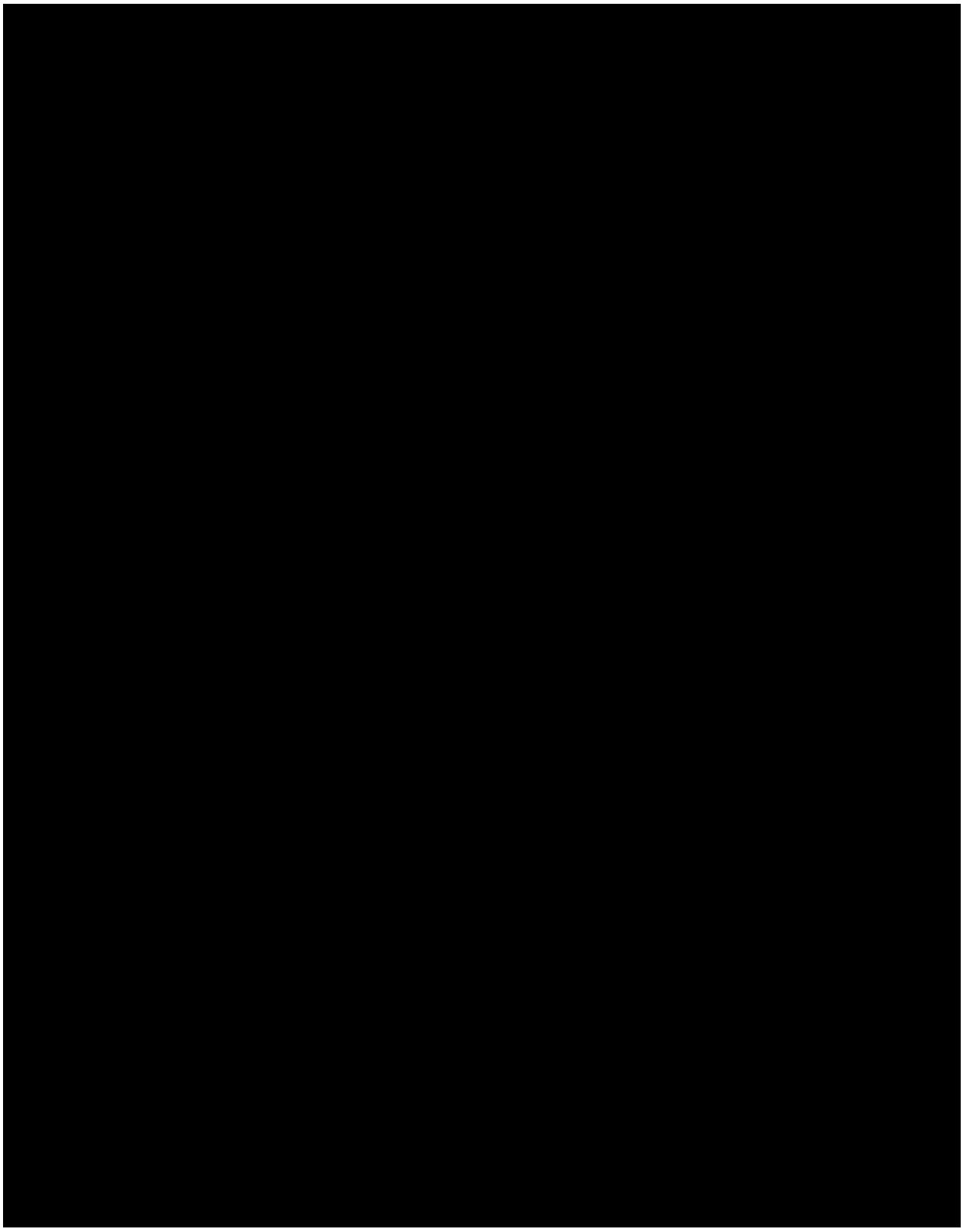


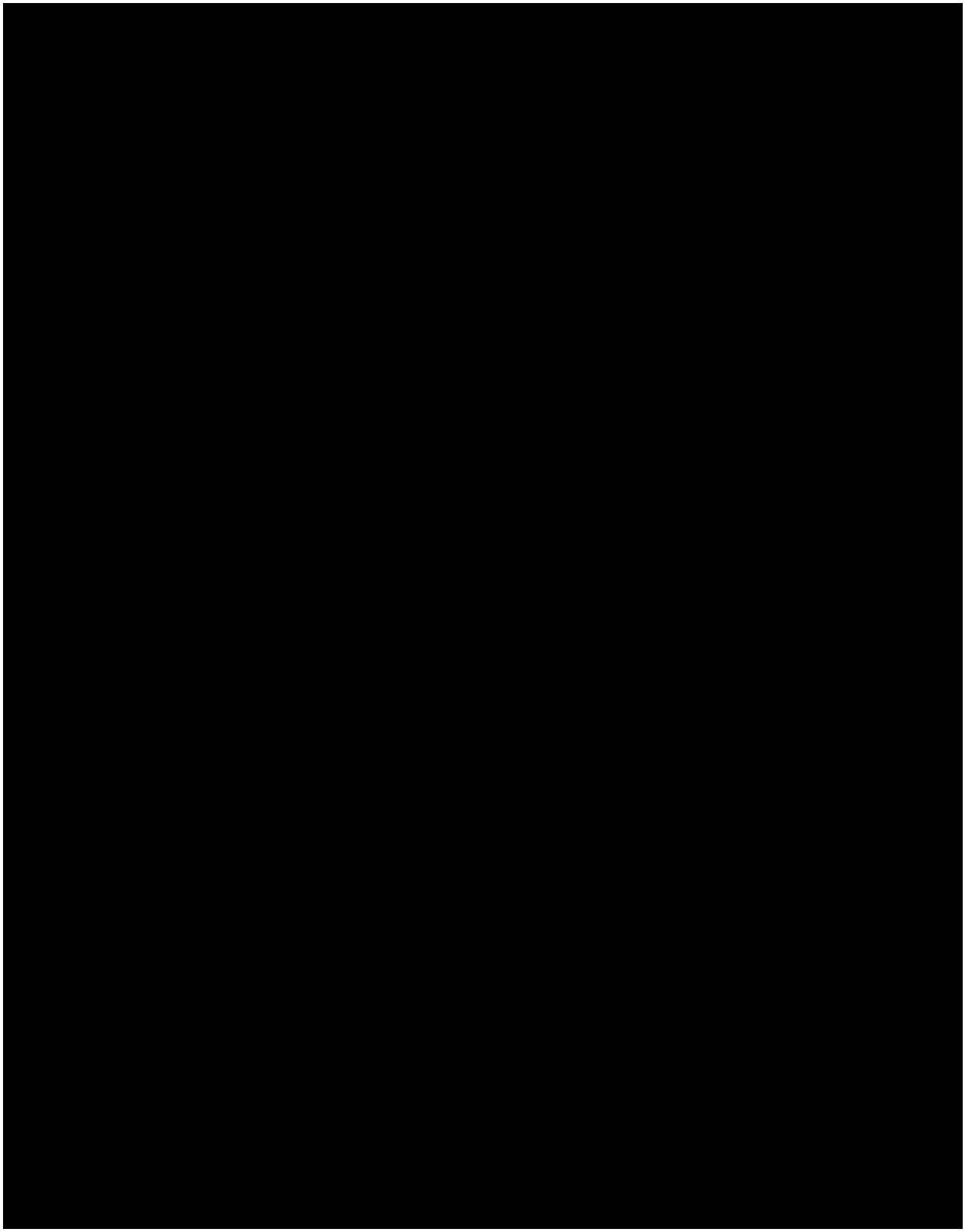


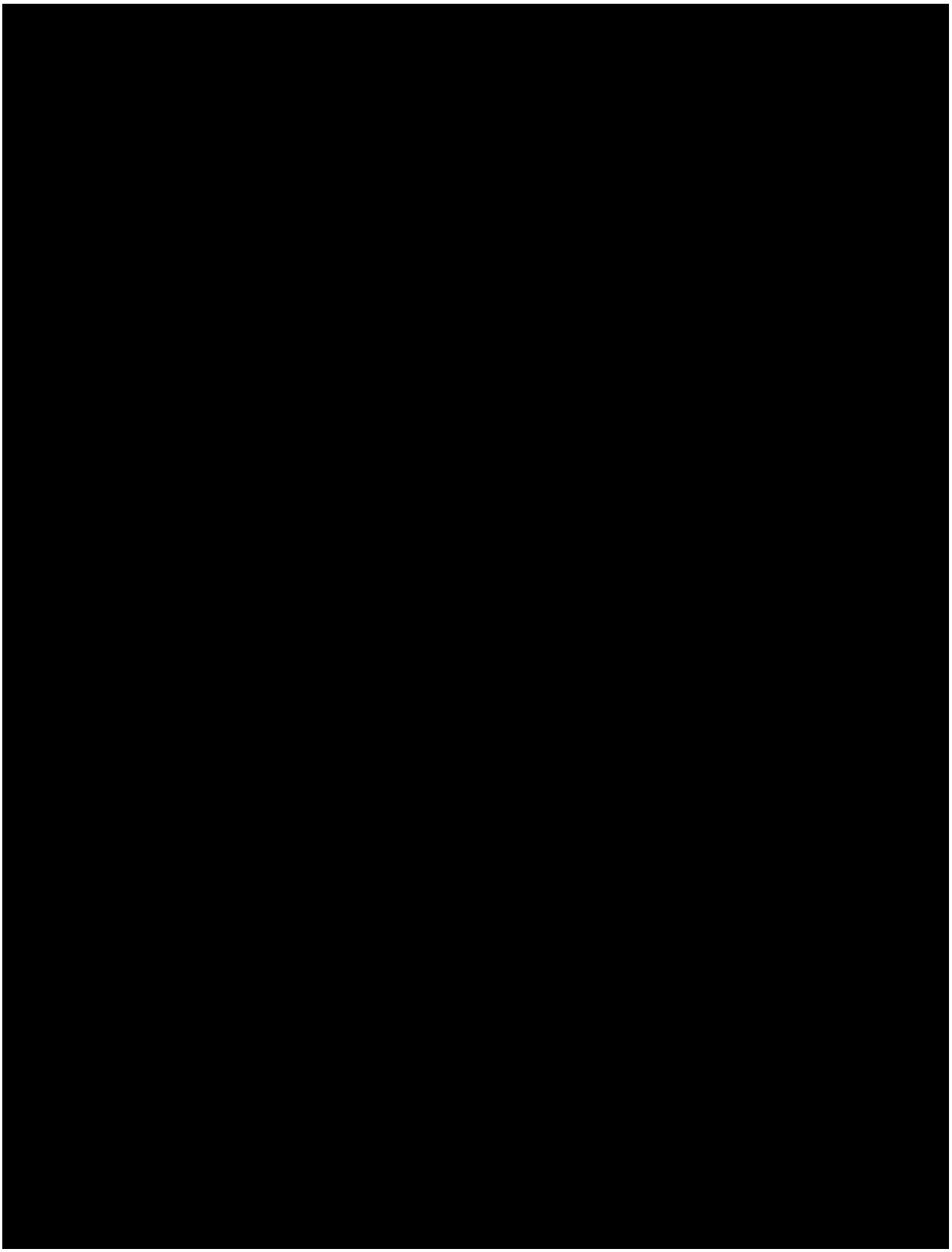












## 6.4 Kaplan-Meier type analyses and graphs

Because this trial has two consecutively recruited trial parts with different allocation ratios that may differ in important predictive covariates and since one treatment group is not even represented in trial part 1, simply pooling data across trial parts to produce an overall Kaplan-Meier graph would not be appropriate. For this reason an alternative approach for producing such a Kaplan-Meier type graph is described in section 6.4.1 below. Such graphs and within any trial part Kaplan-Meier plots will follow the general principles described in section 6.4.2.

Kaplan-Meier type analyses and plots will be provided (including by trial part) for the primary endpoint, key secondary endpoints, secondary endpoints and

### 6.4.1 Overall Kaplan-Meier type graphs across trial parts

A Cox model stratified for treatment group (placebo, 50 mg, 150 mg or 300 mg) and time since index MI (< 6 months, ≥ 6 months), with trial part as a covariate will be used to obtain Breslow estimates (Breslow 1974) for the predicted survival curves for each combination of treatment group, time since index MI and trial part – including for the 50 mg group in trial part 1. These predicted survival curves will then be aggregated for an appropriate cohort of patients within each of the 4 treatment groups to obtain a single overall Kaplan-Meier type graph for the 4 treatment groups.

The use of a Cox model stratified for treatment group (placebo, 50 mg, 150 mg or 300 mg) – and also time since index MI (< 6 months, ≥ 6 months) – is a suggestion by Zhang et al. (Zhang et al. 2007) and Terry Therneau (personal communication, 2011). As pointed out by Zhang et al. such a model makes no assumption of proportional hazards between treatment groups or time since index MI groups. Thus, it is more appropriate for obtaining a graphical representation that allows an inspection of the relative effectiveness of treatment groups over time. Using trial part as a covariate and assuming the same proportional effect of trial part for the whole follow-up period for each of the 8 strata (treatment group / time since index MI)

allows extrapolation of what a 50 mg group in trial part 1 would have been like, as suggested by Terry Therneau (personal communication, 2011). To assess the assumptions involved the predicted survival curves from this model for the placebo, 150 mg and 300 mg groups in each stratum and for the 50 mg group in all trial part 2 strata will be compared graphically to those from a model that treats trial part as an additional stratification variable. Should the assumption be strongly violated with respect to trial part having a constant effect over time, it would be an alternative to assume a trial part effect that is only proportional over specific time intervals (e.g. up to 90 days, day 91 to day 183, days 184 to 1 year, 1 year to 2 years, > 2 years).

The overall aggregated cohort survival estimate will be obtained from the estimates for each treatment group / time since index MI / trial part combination using the Hakulinen method (Hakulinen 1982). Compared with the more simple Ederer method this is expected to more closely reflect what might have been obtained, if this study had been conducted with 4 treatment arms from the start. The reason is that the follow-up time after which administrative censoring at trial end will occur will differ – by design – between the trial parts (Thomsen et al. 1992) (Therneau and Grambsch 2000, pp. 261-287).

The cohort for which the cumulative failure rate is estimated using the Hakulinen method will be constructed so that within each trial part the distribution of time since index MI and distribution of maximal potential follow-up is the same for each treatment group. To achieve this within each trial part

1. All the patients in all treatment groups and time since index MI strata in that trial part are selected.
2. Their length of follow-up $\mathbf{F_i}$ and time since index MI information is retained.
3. These patients are assumed to be the (identical) hypothetical recruited cohorts for each of the four treatment groups.

Note that the length of follow-up $\mathbf{F_i}$ is taken to be

- for those patients that were censored: the time from randomization to the maximum actual follow-up for CV events (not just for the first event)
- for patients that had a non-fatal version of the endpoint of interest (e.g. non-fatal MI in case of analyzing the primary MACE endpoint) the same approach as for those patients that were censored will be used
- for patients that had a fatal version of the endpoint of interest (e.g. fatal MI in case of analyzing the primary MACE endpoint) the maximum potential follow-up will be used, i.e. the time from randomization to the maximal potential follow-up date, defined to be the analysis cut-off date (see Section 2.3.7).

For each patient **i** let $\mathbf{S_i(t)}$ be the estimated survival curve based on that patient's treatment group, time since index MI and trial part. If patients are ordered from shortest to longest possible follow-up time $\mathbf{F_{(1)}, F_{(2)}, \ldots, F_{(K)}}$, then the Hakulinen estimate (Hakulinen 1982) for the cumulative survival function $\mathbf{S(t)}$ is calculated recursively for any $\mathbf{F(k) < t \leq F(k+1)}$ as

$$\mathbf{S(t) = S(F_{(k)}) \times [\ \Sigma_{i=1,\ldots,n}\ S_i(t) \times 1\{\ F_i > t\ \}\ ] / [\ \Sigma_{j=1,\ldots,n}\ S_j(F_{(k)}) \times 1\{\ F_j > F_{(k)}\ \}\ ]}$$

as described by Thomsen et al. (Thomsen et al. 1992), Nielsen (Nielsen 1997) and Therneau and Grambsch (Therneau and Grambsch 2000, pp. 261-287).

Note that valid confidence intervals cannot be obtained for these estimates in a straightforward fashion, because of the artificially inflated "apparent sample size" resulting from the way in which the algorithm above ensures balanced time since index MI/trial part across treatment groups and censoring patterns. If pointwise confidence intervals are required (e.g. primary and secondary endpoints) one could create e.g. B=2,000 bootstrap samples, repeat the whole process above (starting from fitting the Cox model) for each bootstrap sample and then obtain interval percentile (or BCa) Bootstrap confidence intervals as described by Efron and Tibshirani (Efron and Tibshirani 1993, pp. 170-176 and pp. 184-188) e.g. for times t=1 year, 2 years, 3 years and 4 years. algorithm might be applied.

### 6.4.2 Kaplan-Meier graphs for each trial part and general principles

The approach for Kaplan-Meier graphs closely follows the recommendations by Pocock et al. (2002).

Cumulative event rate estimates (1 – estimated survivor function) in % versus time of follow-up in years will be shown based on Kaplan-Meier estimates. Compared to showing the survivor function this maximizes detail without needing a break in the scale.

While "technically, any survival plot can be extended right through to the longest follow-up time" as Pocock et al. point out "this extension is not good statistical practice, since for any such plot the eye is drawn to the right (i.e. where the plot finishes), which is where there is least information and greatest uncertainty". Kaplan-Meier curves will therefore only drawn up to the time at which the total population at risk for the event in question has decreased to 10% of the size of the analysis set. This matches the recommendation by Pocock et al. who wrote that they "[..] recommend that survival plots be halted once the proportion of patients free of an event, but still in follow-up, becomes unduly small. […] What constitutes "unduly small" is open to debate and depends on the context. It will often be reasonable to curtail the plot when only around 10–20% are still in follow-up." As CANTOS is a large outcome trial the lower end of this range (10%) will be used.

For the kind of Kaplan-Meier type graphs described in section 6.4.1 above, this will be taken to mean that the graph will be truncated at the length of follow-up achieved by 10% of patients across all trial arms and trial parts. This means that the graph for the 50 mg dose will be continued somewhat past the point up to which 10% of 50 mg patients were followed up.

The number of patients at risk in each group will be displayed below the x-axis at year=0, 1, 2, 3… and each subsequent year of follow-up to provide information on the extent of follow-up. Pointwise 95% confidence intervals at each full year may be shown if feasible. In line with the recommendations by Pocock et al. this provides some measure of statistical uncertainty and ensures that any visual signs of treatment differences do not look more convincing than they really are. Results of inferential statistics (e.g. primary model-based hazard ratios, 95% confidence intervals and p-value for primary endpoint) will also be shown on the figure. Kaplan-Meier tabulations will show number at risk, number failed, cumulative number failed, cumulative number censored, Kaplan-Meier estimates of failure rate with pointwise 95% confidence intervals. Similar tabulations will be provided by time interval also including the estimated hazard at interval midpoint.

It is recognized that just as discussed by Pocock et al. caution is needed in interpreting the shape of survival plots in particular due to the lack of any pre-specified hypothesis about its shape and the lack of statistical power to explore subtleties of treatment difference other than the overall comparison.

### 6.4.3 Calculation of number needed to treat

The number needed to treat (to benefit) will be calculated using the method of Altman and Andersen (Altman and Andersen 1999) for each dose for different durations of treatment as

$$\mathbf{NNT\ (t) = 1\ /\ [\ S_{placebo}(t)^{\text{hazard ratio for dose versus placebo}} - S_{placebo}(t)\ ]}$$

using S(t) calculated as described in section 6.4.1. It will be extrapolated as described by Altman and Andersen to the 5 year timepoint from the timepoint of follow-up achieved by 10% of patients across all trial arms and trial parts $T_{10\%}$ under the assumption of approximately constant event rates and benefit as

$$\mathbf{NNT\ (5\ years) = (\ T_{10\%}\ /\ 5\ years\ ) \times NNT\ (T_{10\%}).}$$

3 year and 1 year NNTs will be calculated and it is expected that this can be done without such an extrapolation unless the study is stopped an interim analysis. Confidence intervals with appropriate coverage probabilities will be calculated using Bootstrap methods.

## 6.5 Methodology for dose response evaluation

Dose-response estimation will be performed to support the evaluation of a number of endpoints as indicated in the protocol, among them the hazard ratios derived from the analysis of the primary endpoint.

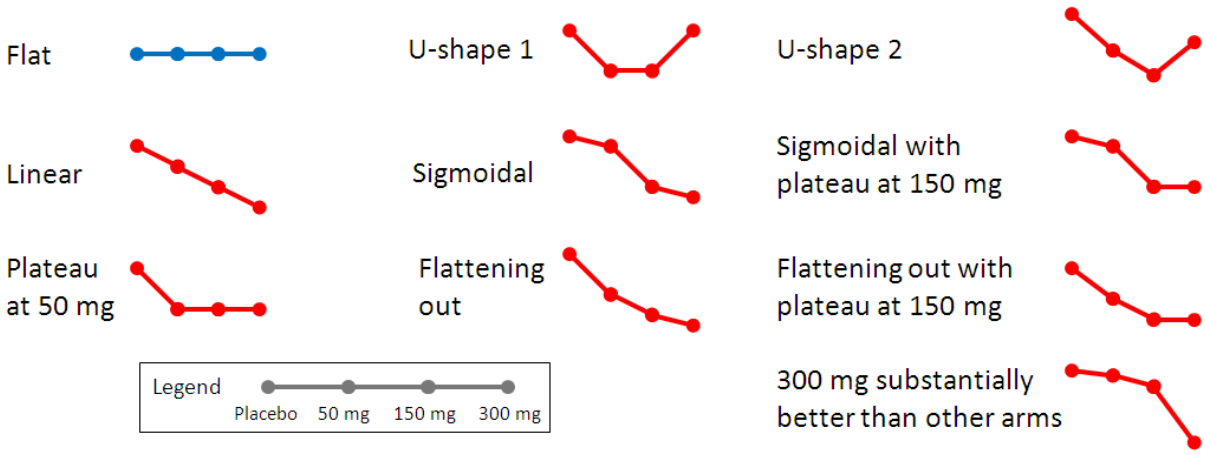
Dose-response estimation will be done using parametric dose-response modelling as described in Pinheiro et al. (2014) combined with model averaging techniques (see Bornkamp et al. 2015). The model averaging approach will use a weighted subset of seven possible parametric models as the estimate of the dose response. The parametric models are a linear model, a quadratic model, a model linear in log dose, an emax model, a sigmoid emax model, an exponential model, and a beta model as implemented in the R package DoseFinding. The model averaged dose response and confidence intervals will be derived from a parametric bootstrap. In addition, comparisons between doses will also be derived with their corresponding confidence intervals.

The time to event endpoints analyzed using the general approach as described in Pinheiro et al. (2014) will use the estimated hazard ratios and covariance matrix derived from a parametric regression model for survival endpoints for which the treatment estimates are given a proportional hazards model. Estimates for the analysis of other endpoints will be derived from fitting models that account for measurements collected over time whenever applicable. [redacted]

The estimation of the dose response will be based on the methods described in Pinheiro et al. 2014 as a more general framework for the derivation of dose responses when the response is any form of derived statistic for the dose groups with their corresponding variance covariance matrix.The derived dose response will be obtained from a model that is the average of at least three of fitted models of dose response where the weights will used the generalized AIC

measure as described in Pinheiro et al 2014. The corresponding mean dose response curve and its confidence intervals will be derived from a parametric boostrap procedure also described in the same reference.

**Figure 6-6 Candidate dose response patterns**




# 7 Safety evaluation

Safety will be evaluated based on the safety set (SAF). The assessment of safety will be based primarily on the assessment of potential and identified risks defined in the risk management plan (RMP), the frequency of adverse events, laboratory abnormalities, and serious adverse events suspected by the investigators to be related to study treatments. Other safety data (like vital signs, ECG) will be summarized as appropriate.

Selected safety analyses will be based on looking at events and variables recorded while patients were still on investigational drug (i.e. data will be considered censored at the latest one quarter year + 28 days = 119 days after the last study injection, and up-to/including the EOS visit of the core phase), as supportive analyses.

Since the study is intended to be the only pivotal trial in the regulatory submission, selected key safety analyses including disposition, demographics, exposure, adverse events, serious adverse events, key outputs for adverse events of special interest, deaths, laboratory abnormalities and laboratory values over time will be repeated for those groups specified in table 2.3-2.

Selected safety analyses will also be shown by time period (from randomization to $\leq$ 3 months post-randomization, > 3 to $\leq$ 6 months, > 6 to $\leq$ 9 months, > 9 to $\leq$ 12 months, > 12 months to the end of core study visit) to evaluate any changes in safety during more short-term and longer term treatment.

## 7.1 Safety assessment approach accounting for multiplicity

No adjustments for multiplicity will be made for safety assessments. Adjustments will be made for multiplicity of the primary and key secondary efficacy endpoints, as described earlier in this document.

## 7.2 Statistical models used in multiple safety contexts

Safety data will be summarized across the two trial parts and the two randomization plans in trial part 2, both of which might influence safety reporting as described in section 2.4.

For this purpose hazard ratios versus placebo will be estimated for binary safety outcomes (e.g. at least one occurrence of adverse events) using Cox regression stratified for time since index MI and trial part. Time to event methods have been chosen over relative risk regression or logistic regression to account for differential follow-up This can be implemented using SAS code similar to the following:

```
proc phreg data=dataset;
  class treatment(ref="Placebo") <time since index MI>
      <trial part>;
  model time*event(0) = treatment / ties=exact risklimits;
 strata <time since index MI > <trial part>;
run;
```

Trial part by treatment interactions will be investigated. For selected events, additional predictive baseline covariates may be added to the model, namely age, sex, race, region and other event specific variables. As summary statistics patients with an event per 100 patient years of follow-up (defined as follow-up to first event or censoring) will be reported.

Rates of multiple occurrences of safety events will be compared between treatment groups for selected safety events of special interest, where such an analysis is of interest. A Negative Binomial regression including terms for treatment, trial part, and log (time at risk) as offset variable will be used. This model was chosen, because the number of recurrent events in a population will follow a negative binomial distribution if there is a distribution of different individual Poisson event rates in the population that follows a gamma distribution (Glynn et al. 1993) (Glynn and Buring 1996). While the population distribution may not exactly follow a gamma distribution, the maximum likelihood estimation of the additional dispersion parameter compared to the Poisson distribution from the data allows the negative binomial model approximate the data in this respect and to capture the extent of unexplained heterogeneity in the population. The logarithm of the time at risk is used as an offset (i.e. as a covariate with coefficient 1), because a log link function is used in order to effectively analyze the rate of events per time unit. The following SAS code illustrates the statistical model and methods intended for this analysis:

```
PROC GENMOD DATA=events;
 CLASS treatment(ref='Placebo') <trial part & time since index MI>;
  MODEL <number of events> = treatment
      <trial part & time since index MI> / OFFSET=<log time at risk>
```

```
        DIST=NB LINK=LOG LRCI;
RUN;
```

As matching summary statistics, events per 100 patient years of follow-up will be reported. For selected events, additional potentially predictive covariates will be added to the model, namely at least age, sex, race and region. When selecting repeated events it has to be ensured that a single adverse event case report form entry is only counted once (e.g. by ensuring only one record per adverse event sequence number is counted), as analysis dataset may contain multiple records for one CRF report to allow identification via e.g. different SMQs.

When such a model is used for specific adverse events of interest and a sufficient number of events to make this display useful is available, these data will also be displayed graphically as events per unit time of follow-up for different time intervals (e.g. first 6 months, 6 months – 1 year, 1 year to 2 years … or smaller time intervals as appropriate) using meaningful categories of time to show time patterns of event occurrence and potential changes in treatment effect over time. Measures of uncertainty will be provided, as well as information on the extent of overdispersion / population heterogeneity.

Changes from baseline in continuous safety parameters will be plotted over time based on both trial parts using repeated measures mixed model. The following SAS code illustrates the statistical model and methods intended for this analysis:

```
proc mixed data=dataset;
  class patient visit treatment <trial part>
      <time since index MI>;
  model <change from baseline> = treatment visit treatment*visit
      baseline <trial part>
      <time since index MI>
      / s ddfm=kr;
  repeated visit / subject = patient type = un;
run;
```

with appropriately transformed values. Comparisons of each dose to placebo at each visit will be conducted if feasible. Additional analyses will add potentially predictive covariates, namely age, sex, race, region, education, exercise level, alcohol use, smoking history, BMI and hypertension to the model. The following variables will be analyzed on the log-scales and results will be back-transformed to report geometric means and ratios of geometric means versus placebo:

- [redacted]
- ALT (SGPT), AST (SGOT), total bilirubin, direct bilirubin, alkaline phosphatase
- Triglycerides, HDL, LDL, VLDL
- hsCRP

All other safety variables will be analyzed without any transformation based on the arithmetic mean and differences in arithmetic means versus placebo.

Panel plots of the results of these analyses showing means/geometric means or differences in means/ratios of geometric means over time will be made available for the clinical study report. 95% pointwise confidence intervals will be provided to allow judgment on the variability in the underlying data, but due to the obvious multiplicity issues consideration of these confidence intervals or p-values from the model described above cannot form the basis of any formal inference.

## 7.3 Other safety evaluation

### 7.3.1 Adverse Events

#### Adverse events between informed consent and randomization

Adverse events between informed consent and randomization will be summarized by primary system organ class and preferred term for all randomized patients (as opposed to the safety set). This summary will include events that would qualify as trial endpoints if they occurred after randomization, but which are to be reported as AEs prior to randomization.

Serious adverse events occurring after informed consent will be reported for all screened patients, as well as separately for all randomized patients and for screen failures.

#### Adverse events (AEs) excluding trial endpoints

The incidence of treatment emergent AEs (events started on the first dose of study medication or events present prior to start of double-blind treatment but increased in severity based on preferred term, up-to/including the earlier date of 119 days after the last study injection or the EOS visit of the core phase) excluding trial endpoints will be summarized by primary system organ class (SOC), preferred term and also by severity and relationship to study treatment. Standardized MedDRA Queries (SMQs) may also be employed. The MedDRA version used for reporting the study will be clearly identified.

Any adverse events reported by the investigator with the onset date after randomization and up-to/including the end of core phase visit and prior to clinical database lock will be included in AE analyses irrespective of how long after last study drug intake. Adverse events occurring after a patient's end of core phase visit will be reported as part of the washout phase/long-term safety extension phase.

If a patient reported more than one AE with the same preferred term, the AE with the greatest severity will be presented. If a patient reported more than one AE within the same primary system organ class, the patient will be counted only once with the greatest severity at the system organ class level, where applicable.

The number and percentage of patients reporting any AE during the double-blind phase of the study will be summarized by primary system organ class, preferred term and treatment. The most common adverse events reported (reported by ≥ 2 % of patients in any group for each preferred term or ≥ 2 % in any group for each SMQ) table will be presented in descending frequency starting from the most common event.

Separate summaries will be provided for AEs suspected by the investigator to be related to study drug, deaths, SAEs, and AEs leading to discontinuation.

### Events reported on endpoint pages

The following events reported by investigators on the trial endpoint pages will be summarized by treatment group for safety purposes, because they are as per the study protocol not reported as adverse events (further sub-categorizations may be added as needed):

- All-cause death
  - Non-cardiovascular death
    - Non-cardiovascular death due to accident/trauma
    - Non-cardiovascular death due to respiratory failure
    - Non-cardiovascular death due to infection
    - Non-cardiovascular death due to sepsis
    - Non-cardiovascular death due to malignancy
    - Non-cardiovascular death due to suicide
    - Non-cardiovascular death due to renal failure
    - Non-cardiovascular death due to liver failure
    - Non-cardiovascular death due to other non-vascular reason
  - Cardiovascular death
    - Coronary Heart Disease death
      - Cardiovascular death due to myocardial infarction
      - Cardiovascular death due to congestive heart failure/cardiogenic shock
      - Cardiovascular death directly related to revascularization
      - Cardiovascular death due to arrhythmias
      - Cardiovascular death due to witnessed sudden death
      - Cardiovascular death due to unwitnessed sudden death seen within 24 hours
      - Cardiovascular death (unwitnessed death)
      - Cardiovascular death (death due to unknown cause)
      - Cardiovascular death due to other vascular causes identified as Ischemic Heart Disease (i.e. fulfilling narrow “Ischaemic heart disease” SMQ)
    - Cardiovascular death due to other vascular causes not identified as Ischemic Heart Disease (i.e. not fulfilling narrow “Ischaemic heart disease” SMQ)
    - Cardiovascular death due to atherosclerotic non-coronary vascular disease
    - [redacted]
    - Cardiovascular death due to stroke
      - Cardiovascular death due to non-hemorrhagic stroke
      - Cardiovascular death due to intracranial hemorrhage
- Stroke or TIA
  - Stroke
    - Non-hemorrhagic cerebral infarction

    - Non-hemorrhagic infarction with hemorrhagic conversion
    - Primary hemorrhagic intracerebral stroke
    - Stroke of uncertain type
  - Transient Ischemic Attack requiring/not requiring hospitalization
- Coronary vessels with >50% stenosis
  - Coronary vessels >50% stenosis: Left Main
  - Coronary vessels >50% stenosis: LAD
  - Coronary vessels >50% stenosis: LCX
  - Coronary vessels >50% stenosis: RCA
  - Coronary vessels >50% stenosis: Bypass Graft
- Myocardial infarction (with 12 sub-classifications as below)
  - Spontaneous myocardial infarction with/without new ST segment elevation with/without new Q waves identified
  - PCI-related myocardial infarction with/without new ST segment elevation with/without new Q waves identified
  - CABG-related myocardial infarction with/without new ST segment elevation with/without new Q waves identified
- Unstable Angina (with 4 sub-classifications as below)
  - Unstable angina requiring/not requiring unplanned revascularization and requiring/not requiring hospitalization
- Coronary revascularization (with 4 sub-classifications as below)
  - Urgent/Elective percutaneous coronary intervention without stenting/using drug eluding stent/bare metal stent/unknown type of stent
  - Urgent/Elective coronary artery bypass graft (1 bypass graft/>=2 bypass grafts)

- Critical limb ischemia (with 16 sub-classifications as below)
  - Critical limb ischemia requiring hospitalization
    - Critical limb ischemia in left lower extremity/left upper extremity/right lower extremity/right upper extremity requiring revascularization/not requiring revascularization and requiring hospitalization
  - Critical limb ischemia not requiring hospitalization
    - Critical limb ischemia in left lower extremity/left upper extremity/right lower extremity/right upper extremity requiring revascularization/not requiring revascularization and not requiring hospitalization
- Limb amputation due to vascular cause (with 10 sub-classifications as below)
  - Amputation of right lower extremity/right above knee amputation/…/left upper extremity

- New onset diabetes
  - New onset diabetes based on start of new diabetes medication
  - New onset diabetes based on laboratory criteria
    - New onset diabetes based on HbA1c only
    - New onset diabetes based on FPG only
    - New onset diabetes based on HbA1c and FPG

### 7.3.2 Laboratory data

Laboratory values that the laboratory reports to be below or above the limit of quantification will be imputed as 0.5 × and 1.5 × the respective limit of quantification. Should there be a substantial proportion of values below the lower limit of quantification and/or above the upper limit of quantification for an important laboratory parameter (see Table 7-1), the summary statistics (mean, standard deviation) will also be calculated as the maximum likelihood estimates using a parametric model for data that can be right censored and left censored at the upper and lower limit of quantification, respectively, using the following methods

- Arithmetic mean and SD are calculated using PROC LIFEREG assuming a normal distribution

- Geometric mean and CV%(geometric mean) are calculated using PROC LIFEREG assuming a log-normal distribution and back-transformed
- Empirical median will not be calculated in the case of censoring. Empirical minimum will not be calculated if there are values < lower limit of quantification, similarly the empirical maximum if there are values > upper limit of quantification.
- Quartiles from the parametric distribution can also be calculated upon request when Q1, median or Q3 does not lie between the lower and upper limits of quantification.

The summary of laboratory evaluations will be presented for three groups of laboratory tests (Hematology, Serum chemistry and Urinalysis). Summaries will consist of descriptive tables stratified by treatment and time point.

Descriptive summary statistics (mean, median, standard deviation, Min and Max) for the change from baseline to each study visit will be presented. These descriptive summaries will be presented by laboratory test and treatment group. Change from baseline will only be summarized for patients with both baseline and post baseline values and will be calculated as:

change from baseline = post baseline values – baseline value

In addition, shift tables will be provided for all parameters except for creatinine clearance in order to compare a patient's baseline laboratory evaluation relative to the extreme post-baseline value. For the shift tables, the normal laboratory ranges will be used to evaluate whether an extreme post-baseline value was normal, low, or high relative to whether or not the baseline value was normal, low, or high. These summaries will be presented by laboratory test and treatment group.

The frequency and percentage of patients with clinically notable laboratory results after baseline will be tabulated. Clinically notable laboratory results, for those parameters where ranges are available, are given in Table 7-1 below and are based on the FDA Division of Neuropharmacology guidelines. Only patients with laboratory results not notably abnormal at baseline from the central laboratory at baseline will be included in the tabulations.

**Table 7-1 Clinical notable criteria for selected laboratory tests**

| Laboratory parameter (unit) | Lower bound of clinically notable range | Upper bound of clinically notable range |
|---|---|---|
| Hematology | | |
| Hematocrit (%) | 30 | 60 |
| Hemoglobin (g/dL) | 10 | 20 |
| Platelet count (x $10^9$/L) | 100 | 600 |
| RBC (x $10^{12}$/L) | 3.3 | 6.8 |
| WBC (x$10^9$/L) | 3.0 | 15.0 |
| Basophils (%) | - | 6 |
| Eosinophils (%) | - | 10 |
| Lymphocytes (%) | 10 | 60 |
| Monocytes (%) | - | 20 |
| Absolute neutrophils (x $10^9$/L) | 1 | 12 |
| Chemistry | | |

| Laboratory parameter (unit) | Lower bound of clinically notable range | Upper bound of clinically notable range |
|---|---|---|
| Albumin (g/L) | 25 | 60 |
| Alkaline Phosphatase (U/L) | - | 280 |
| BUN (mmol/L) | 0.7 | 14.3 |
| Calcium (mmol/L) | 1.87 | 2.89 |
| Chloride (mmol/L) | 85 | 119 |
| Creatinine (mcmol/L) | 18 | 221 |
| LDH (U/L) | 0 | 500 |
| Potassium (mmol/L) | 3 | 6 |
| AST (U/L) | - | 100 |
| ALT (U/L) | - | 110 |
| Total Bilirubin (mcmol/L) | - | 43 |
| Total Protein (g/L) | 40 | 95 |
| Uric Acid (mcmol/L) | 89 | 595 |
| BUN/Serum Urea (mmol/L) | 0.7 | 14.3 |
| Sodium (mmol/L) | 125 | 154 |

Patients' worst CTC classifications as per table 7-2 at any point during the study will also be provided. The number of patients with eGFR declines compared to baseline of ≥ 25%, ≥ 40% or ≥ 50%, or by ≥ 30 ml/min/1.73 $m^2$ at any point in the study will be summarized as well as those with serum creatinine increases compared to baseline of ≥ 50%, ≥ 0.5 mg/dL, ≥ 2.0 mg/dL, ≥ 2.5 mg/dL or ≥ 3.0 mg/dL.

[redacted]

Additionally, the occurrence of nephropathy based on selected adverse event groupings as defined in the CRS, doubling of post-baseline serum creatinine in comparison to baseline, endstage renal disease (ESRD) defined as dialysis, renal transplant, or serum creatinine >530 µmol/L (6.0 mg/dL) and time to non-cardiovascular death due to renal failure will also be analyzed. An overall composite will also be provided. The continuous log-transformed urine albumin creatinine ratio will be also analyzed as well as eGFR over time and abnormalities will be considered.

**Table 7-2 NIH CTC version 4.0 grades for Chronic kidney disease, which can be assessed in this trial**

| CTC terms | Grade 1 | Grade 2 | Grade 3 | Grade 4 |
|---|---|---|---|---|
| Chronic kidney disease | eGFR† <LLN - 60 ml/min/1.73 $m^2$ | eGFR† 59 - 30 ml/min/1.73 $m^2$ | eGFR† 29 - 15 ml/min/1.73 $m^2$ | eGFR† <15 ml/min/1.73 $m^2$; dialysis or renal transplant indicated* |

† CrCl not assessed in this study

* in this study to be defined as report of kidney transplant or dialysis as non-drug therapy

### 7.3.3 Vital signs

Descriptive summary statistics (mean, median, standard deviation, Min, Max) for the change from baseline to each post baseline visit will be presented. These descriptive summaries will be presented by vital sign and treatment group. Change from baseline will only be summarized for patients with both baseline and post-baseline values and will be calculated as:

**change from baseline = post-baseline value – baseline value**

The frequency and percentage of patients with clinically notable vital signs (based on the worst value in changes from baseline) will be tabulated. Clinically notable vital sign results are provided in Table 7-3 below. Pulse pressure will be derived as systolic – diastolic blood pressure and reported.

**Table 7-3 Clinically notable changes in vital signs**

| Vital Sign (unit) | Clinically notable criteria |
|---|---|
| Weight (kg) | increase > 10% from Baseline |
| Systolic blood pressure (mmHg) | ≤90 and decrease from baseline of ≥30 or <75 |
| | ≥180 and increase from baseline of ≥30 or >200 |
| Diastolic blood pressure (mmHg) | ≤50 and decrease from baseline of ≥20 or <40 |
| | ≥105 and increase from baseline of ≥20 or >115 |
| Pulse (bpm) | ≤50 and decrease from baseline of >30 or <40 |
| | ≥120 and increase from baseline of >25 or >130 |

### 7.3.4 Electrocardiogram (ECG)

In the CANTOS trial, baseline and yearly thereafter ECGs will be taken; ECG results will be centrally read for changes including QTc changes and silent MIs for clinical CAD endpoints.

The following quantitative variables will be summarized at each visit using standard summary statistics: ventricular rate (heart rate), RR interval, PR interval, QRS duration, QT interval, and corrected QT interval (QTcF/QTcB). The difference between treatment groups in observed values and changes from baseline with associated 95% confidence interval will be provided for each of these criteria. QTc data will be analyzed for both the Fridericia (primary) and Bazzett's (secondary) corrections.

Additionally, the number and proportion of patients with

- QT > 550 ms
- QT > 500 ms
- QTc > 500 ms
- QTc > 480 ms
- QTc > 450 ms
- QTc changes from baseline > 30 ms
- QTc changes from baseline > 60 ms
- PR > 250 s

will be summarized .

In addition, shift tables comparing baseline ECG results (normal, abnormal, not available, total) with the maximum on-study result (normal, abnormal, not available, total) will be provided for each variable.

Newly occurring or worsening abnormalities will also be summarized by treatment group. Results in elderly patients (≥ 65), by gender, in patients with congestive heart failure, elevated baseline QT/QTc intervals (for QT/QTc analyses) and baseline HR <55 bpm will be presented.

## 7.4 Safety topics of interest

This section outlines the currently planned analyses based on the current safety profiling plan. The safety profiling plan is an evolving document outlining the clinical safety plan at any stage of the clinical development of a compound. Thus, due to the length of the trial the analyses based on potential and identified risks may have to be updated according to new information while the trial is ongoing.

In general, analyses for adverse events of special interest based on a list of search criteria stored in the Novartis Clinical REsearch Documentation and Information system (CREDI). For each DMC any analyses to be included in DMC reports will be communicated to the academic independent statistician. Number of affected subjects and percentage will be provided by treatment arm. Risk ratio and 95% confidence interval will be provided for each dose versus placebo. If unbalances occur, time to first occurrence and number of occurrences will be evaluated, if appropriate and not already pre-specified. Dependent on the risk medical history or baseline characteristics may have to be considered further beyond what is prespecified below.

Table 7-4 summarizes the AEs that will be included in any summaries and analyses of AEs of special interest.

Frequency tables will be presented for each AEs of interest as well as the associated MedDRA Preferred Terms. Additionally, Kaplan-Meier plot will be presented for time-to-first event for each AEs of interest unless otherwise specified.

**Table 7-4 List of safety topics of interest**

| Safety topic |
|---|
| Infections/opportunistic infections |
| Neutropenia |
| Thrombocytopenia |
| [redacted] |
| Malignancies |
| Hepatic safety |
| Autoimmunity |
| Injection site reactions |
| Disorders of lipoprotein metabolism |
| Vertigo/dizziness |
| Hypoglycemia |

### 7.4.1 Infections / opportunistic infections

Infections as adjudicated will be summarized based on the adjudicated information including whether the infection was classified as opportunistic, whether the infection was typical or atypical based on the patient's medical history, response to standard antibiotic therapy, whether it was a chronic recurrent infection/exacerbation of a previous infection and the presence of risk factors for infections.

The analyses below will be conducted separately for each of the followings if feasible

- Adjudication committee confirmed infections
- All investigator reported AEs of infection (independent of adjudication)
- All investigator reported SAEs of infection (independent of adjudication)

Additionally, all investigator reported infections will be identified using a list of search criteria in the CRS which will be stored in CREDI. The occurrence of such events will be reported by MedDRA preferred term.

These will be classified by severity, seriousness, whether intravenous antibiotics were required (report of use of i.v. antibiotic as a concomitant medication), whether infection AEs/SAEs were suspected to be related to study drug by the investigator and whether study drug was discontinued due to infections. Non-cardiovascular death due to infection/sepsis will also be analyzed. Time to first infection will also be analyzed both in a stratified Cox regression model, as well as using Kaplan-Meier type analyses.

Since infections are expected to be sufficiently frequent in a long-term study to allow the use of multiple covariates, age, sex, race, and region will be added to the Cox regression models stratified by time to index MI and trial part.

The following analyses will also be presented by time period (from randomization to $\leq$ 3 months post-randomization, > 3 to $\leq$ 6 months, > 6 to $\leq$ 9 months, > 9 to $\leq$ 12 months, > 12 to $\leq$ 18 months, > 18 to $\leq$ 24 months, > 24 to $\leq$ 30 months and so on in 6 month intervals for as long as at least 15% of the trial population are still at risk:

- The number of patients at risk, the number of patients with the specified event, and %
- Annualized event rate (AER) with associated 95% CI. The annualized event rate (AER) is calculated as the total number of patients with specific event divided by the cumulative exposure of time to the first specified events.

Rates of infections including repeated infections will also be presented graphically using a non-parametric estimator of the mean cumulative function, analogous to the Nelson-Aalen estimator of the cumulative hazard function (Johnston and So (2003)) and will be summarized including graphs of infections in different time periods. The following time intervals will be used: from randomization to $\leq$ 3 months post-randomization, > 3 to $\leq$ 6 months, > 6 to $\leq$ 9 months, > 9 to $\leq$ 12 months, > 12 to $\leq$ 18 months, > 18 to $\leq$ 24 months, > 24 to $\leq$ 30 months and so on in 6 month intervals for as long as at least 15% of the trial population are still at risk.

Inferential analysis of the rate of infections will use a negative binomial regression model as described in section 7.2. It will be ascertained by the use of time-dependent period variables whether any potential treatment differences become larger over time, as well as testing for the

interaction of log(time) and treatment. This will be done using a negative binomial model with SAS code similar to the following or equivalent code in another programming language such as R:

```
PROC GENMOD DATA=events;
 CLASS treatment(ref='Placebo') <trial part & time since index MI>
      <time period> sex race region;
  MODEL <number of events> =
      treatment <time period> treatment*<time period>
      <trial part & time since index MI>
      age sex race region / DIST=NB LINK=LOG LRCI
      OFFSET=<log time at risk in time period> ;
  REPEATED SUBJECT=patient / TYPE=UNSTR WITHINSUBJECT=<time period>;
RUN;
```

Time to first infection will be analyzed by each of eight subgroups that might be at higher risk (e.g. elderly patients, diabetic status, patients with low baseline neutrophil counts, current smokers history, obese patients, region, asthma patients, and COPD patients), see definitions in Table 7-5. This will be done in separate stratified cox models for each subgroup, with treatment, subgroup, and treatment*subgroup as covariates.

**Table 7-5 Higher Risk Subgroups**

| | |
|---|---|
| Elderly | >=65 years (yes/no) |
| Diabetics | based on diabetics definition in Table 2.3-1 (yes/no) |
| Low baseline neutrophils | CTC >=3 at baseline (yes/no) |
| Current smokers | Current smokers (yes/no) |
| Obese patients | BMI >=35kg/$m^2$ at baseline (yes/no) |
| Regions | The same as above in Table 2.3-1 |
| Asthma patients | PT=”Asthma” in MH (yes/no) |
| COPD patients | PT=”COPD” in MH (yes/no) |

It will be explored whether any potential risk of infections changes based on whether patients have reported a vaccination within 3 months from/including baseline (yes/no), by modelling the time to first infection including treatment, subgroup (vaccination within 3 months, yes/no), and treatment*subgroup. This analysis will also be done for the 6 month time frame (yes/no - live vaccination within 6 months from/including baseline).It should be noted that live vaccinations within 3 months were not permitted in this study.

Separate summaries of infections and tuberculosis related adverse events (defined by HLTs "Tuberculous infections" and "Mycobacteria identification and serology") will be produced for the patients who had tuberculosis, latent or suspected tuberculosis at the start of the study and received randomized study medication. A patients' listing will be provided.

### Duration of infections

To investigate whether the duration of infections is affected by canakinumab, duration will be summarized for each of the endpoints below:

- Mean of the durations in days for each patient
- Maximum of the durations in days for each patient
- Proportion of total durations of infections over total duration of follow-up for each patient

## 7.4.2 Hepatic safety data

Presentation of hepatic safety data will follow the internal Novartis standards for hepatic safety analyses. The risk of liver toxicities will be quantified and described in terms of onset, dose, duration and severity.

Hepatic effects (LFT abnormality) will be identified with selected adverse events (as defined in the CRS), and the laboratory parameters mainly AST (aspartate aminotransferase; also known as SGOT), ALT (alanine aminotransferase; also known as SGPT), ALP (alkaline phosphatase) and TBL (total bilirubin; conjugated (direct) (DBL) and unconjugated (indirect) bilirubin).

All analyses (whether e.g. Cox regression, analysis of covariance or ordinal regression) will take into account baseline log ALT, log AST, log TBL, log direct bilirubin values (only if available in all patients), alcohol use, age, sex, race and region as long as the number of events allows it.

All analyses, when possible, will be performed with baseline glycemic status as a subgroup and including a treatment by baseline subgroup interaction and repeated with abnormal LFTs at baseline as a subgroup including a treatment by baseline subgroup interaction.

### Criterion-based "event" tables

Threshold values of interest for liver function tests are given below.

**Table 7-6 Hepatic events**

| Parameter | Criterion |
|---|---|
| ALT | >3xULN; >5xULN; >8xULN; >10xULN |
| AST | >3xULN; >5xULN; >8xULN; >10xULN |
| ALT or AST | >3xULN; >5xULN; >8xULN; >10xULN |
| TBL | >1.5xULN; >2xULN |
| DBL | >1.5xULN; >2xULN |
| TBL and DBL | TBL>1.5xULN+(DBL>ULN,>1.5xULN or >2xULN) |
| | TBL>2x ULN+(DBL>ULN, >1.5xULN or >2xULN) |
| ALP | >2xULN; >3xULN |
| (ALT or AST) & (TBL or DBL) | ALT or AST >3xULN & TBL or DBL >2xULN |
| | ALT or AST >5xULN & TBL or DBL >2xULN |
| | ALT or AST >10xULN & TBL or DBL >2xULN |

| Parameter | Criterion |
|---|---|
| ALP & (TBL or DBL) | ALP >3xULN & TBL or DBL >2xULN |
| (ALT or AST) & (TBL or DBL) & ALP | AT >3xULN & TBL or DBL > 2xULN & ALP < 2xULN (potential Hy's Law case)<br>Note: elevated ALP may suggest obstruction as a consequence of gall bladder or bile duct disease; ALP may also be increased in malignancy. FDA therefore terms Hy's Law cases as indicators of pure hepatocellular injury. This does not mean that cases of AT >3xULN & TBL >2xULN & ALP ≥ 2xULN may not result in severe drug-induced liver injury (DILI). |

AST = Aspartate aminotransferase; also known as SGOT, ALT = Alanine aminotransferase; also known as SGPT, ALP = Alkaline phosphatase, TBL = Total bilirubin, DBL = conjugated (direct) bilirubin

The laboratory results meeting specified criteria standard table will be used to provide the number and percentage of patients having AST, ALT > 3, 5, 8, 10 x ULN or TBL >1.5, 2 x ULN or ALP > 2, 3 x ULN.

The number and percentage of potential Hy's Law cases will be presented by treatment group. Potential Hy's Law cases are defined as those patients with AST, ALT >3xULN & TBL >2xULN & ALP <2xULN at the same lab measurement.

### Shift-tables

A cross-tabulation of baseline and worst post-baseline values by below, within and above normal range categories will be provided. Shift tables will be provided for the parameters AST, ALT, TBL, direct bilirubin and ALP. These summaries will be presented by laboratory test and treatment group.

### Standard laboratory change from baseline tables

The lab standard tables will be produced on all the laboratory parameters including changes from baseline in liver parameters. Continuous liver function test values will be plotted over time and analyses of covariance adjusting for covariates will also be performed.

### Adverse event tables

Selected groupings (e.g., SMQ, HLGT, HLT) of adverse events, as defined in the CRS will be presented including the respective preferred term frequencies.

Notable events may be defined as liver-related death (Non-cardiovascular death due to liver failure); liver transplantation (from the SAEs and AEs); liver-related study drug discontinuation (from the AEs); any of the 3 events.

The relative risk of the selected groupings (e.g., SMQ, HLT, HLGT) on each of the active treatment dosages relative to control will be calculated with 95% confidence intervals.

The number and percentage of patients with selected groupings of adverse events will be presented.

### Graphical displays

For all patients with liver values (AST, ALT, ALP, TBL) matching the thresholds of 5xULN and Hy's Law (Table 7-6), single patient LFT profile graphs will be generated showing all AST, ALT, ALP, TBL lab values reported for this study and the time of study treatment for these patients. An evaluation of Drug Induced Serious Hepatotoxicity (eDISH) plot will also be created.

## 7.4.3 Malignancy

Occurrence of malignancies as adjudicated by the blinded adjudication committee will be summarized based on whether they are

- Primary type (by MedDRA high level group term/high level term including secondary paths and preferred terms)
- Newly detected or represent an existing malignancy
- Whether the patient's history includes this malignancy
- Evidence of metastases and their stage
- Whether the presentation was considered typical or atypical
- Whether and which risk factors there were
- Whether the malignancy was considered unusual and was the response to therapy as expected.

Malignancies from adverse event reports will be identified using the list of search criteria in the CRS. They will be reported as part of the safety analyses described in section 7.3.1. Preferred terms will be reported to allow assessment of the occurrence of malignancies by site.

The analyses discussed below will be performed separately based on the CEC confirmed malignancies, the investigator reported AEs of malignancies, and the investigator reported SAEs of malignancies, if feasible. Time-to-event analyses will be conducted and Kaplan-Meier type plots will be presented. This will also be done for non-cardiovascular deaths due to malignancy.

To assess whether there is an imbalance when follow-up becomes longer, time-to-first event will be explored looking at <= 2 years of follow-up versus >2 years of follow-up, separately. This will be explored in a Cox regression model including treatment, follow-up (<= 2 years or > 2 years) and the interaction term as covariates.

Following analyses will also be presented by time period (from randomization to $\leq$ 3 months post-randomization, > 3 to $\leq$ 6 months, > 6 to $\leq$ 9 months, > 9 to $\leq$ 12 months, > 12 to $\leq$ 18 months, > 18 to $\leq$ 24 months, > 24 to $\leq$ 30 months and so on in 6 month intervals for as long as at least 15% of the trial population are still at risk.

- The number of patients at risk, the number of patients with the specified event, and %
- Annualized event rate (AER) with associated 95% CI. The annualized event rate (AER) is calculated as the total number of patients with specific event divided by the cumulative exposure of time to the first specified events.

### 7.4.4 Thrombocytopenia

Effects with respect to thrombocytopenia will be assessed by the laboratory analyses described in section 7.3.2, additionally criteria for decreased platelet counts per CTC grades as shown in table 7-7 will be reported and baseline platelet counts will be taken into account in this analysis. Shifts versus baseline will also be shown per CTC grade.

**Table 7-7 NIH CTC version 4.0 grades for platelets**

| CTC terms | Grade 1 | Grade 2 | Grade 3 | Grade 4 |
|---|---|---|---|---|
| Platelet count decreased | <LLN-75,000/$mm^3$; <LLN-75.0 x $10^9$/L | <75,000-50,000/$mm^3$; <75.0 - 50.0 x $10^9$/L | <50,000 - 25,000/ $mm^3$; <50.0 - 25.0 x $10^9$/L | <25,000/$mm^3$; <25.0 x $10^9$/L |

The occurrence of the groupings of adverse events will be analyzed using Cox regression as described in section 7.2 also adding the baseline platelet count to the model and looking for an interaction with treatment.

The baseline use of antithrombotic medications will be evaluated as a subgroup and treatment by subgroup interactions will be evaluated for both the occurrence thrombocytopenia (AE of special interest) and bleeding events (Haemorrhages SMQ), using Cox regression.

Time to the 1st-occurrence of CTC grades ≥3, and time to the 1st-occurrence of bleeding events will be plotted over time using Kaplan-Meier type graphs. A negative binomial regression, as described in section 7.2, will be used to analyze the occurrence of the bleeding events over time. A repeated measures analysis will be performed on the continuous platelet counts change from baseline with the following factors/covariates: treatment, time, time by treatment interaction, time since index MI, randomization plan, and baseline platelet count. Platelet counts change from baseline will also be plotted over time.

### 7.4.5 Neutropenia

Drug effects on neutrophils will be assessed by the laboratory analyses described in section 7.3.2, additionally decreased neutrophil counts (occurrence of neutropenia) per CTC grades as shown in table 7-8 will be reported and baseline neutrophil / white blood cell counts will be taken into account in this analysis. Shifts versus baseline will also be shown per CTC grade.

Whether baseline neutropenia or early drops in neutrophil counts (or neutropenia) are indicative of infections (defined using adjudicated events of infection, investigator reported infections, and AEs of special interest) will be explored in two separate Cox regression models; one including treatment, baseline neutropenia and the interaction term as covariates, and the second model including treatment, early drops in neutrophil counts (or neutropenia) and the interaction term as a covariate. There are two definitions for the early drop in neutrophil counts (or neutropenia) subgroup, one using an ordinal variable and one using percentage change from baseline, and these are both computed at visits 3 and 4, see Table 2.3-3. There will therefore be four models for this endpoint and these will be displayed in the same table.

**Table 7-8 NIH CTC version 4.0 grades for neutrophils**

| CTC terms | Grade 1 | Grade 2 | Grade 3 | Grade 4 |
|---|---|---|---|---|
| Neutrophil count | <LLN - 1500/$mm^3$; | <1500 - 1000/$mm^3$; | <1000 - 500/$mm^3$; | <500/mm3; <0.5 x |

| CTC terms | Grade 1 | Grade 2 | Grade 3 | Grade 4 |
|---|---|---|---|---|
| decreased | <LLN - 1.5 x $10^9$/L | <1.5 - 1.0 x $10^9$/L | <1.0 - 0.5 x $10^9$/L | $10^9$/L |

Whether neutrophil differences change over time will be assessed graphically, as well as using a repeated measures analysis of the continuous neutrophil count change from baseline with the covariates: treatment, time, time by treatment interaction, time since index MI, randomization plan, and baseline neutrophil count. Whether the occurrence of abnormalities (especially CTC grade ≥ 3) depends on the potential risk factor baseline neutrophil counts will be investigated using a cox regression model with time to first occurrence of abnormalities (CTC grade ≥ 3) with treatment, baseline neutrophil counts and treatment by baseline neutrophil counts interaction.

The number of patients that discontinued treatment and then returned to normal values or did not return to normal values, as well as those that did not discontinue and then returned to normal values or did not return to normal values will be summarized. The usage of concomitant medications given to normalize white blood cells counts/neutrophil counts will be summarized.

### 7.4.7 Disorders of lipoprotein metabolism

Drug effects on lipids will be assessed by the laboratory analyses described in section 7.3.2. Additional subgroups based on the use of lipid-lowering medications (including classes of these such as statins and fibrates) at baseline will be investigated. CV outcomes are the primary endpoint of this study and sub-group analyses for CV outcomes by baseline lipid values are planned (see section 2.2). Additionally, the dose of statins used will be tracked as described in section 5. The incidence of adverse events such as myopathy that could result from different statin usage patterns may be explored based on the findings. The number of patients with persistently elevated triglycerides (>50% of visits with TG > 500 mg/dL), as well as the occurrence of adverse events such as pancreatitis (acute pancreatitis SMQ) that could result from persistent elevations will be summarized.

### 7.4.8 Hypoglycemia

As per the draft FDA requirements for the evaluation of prevention of diabetes, rates of hypoglycemia will be compared statistically between groups. Hypoglycemic events will be identified using the the SMQ Hypoglycaemia. Comparisons between treatment groups will be both done as time to first event analyses (including reporting the number of affected patients), as well as analyzing count data (number of patients with a certain number of hypoglycemic episodes – i.e. none, 1 episode, 2 episodes etc.) using negative binomial regression as described in section 7.2. Baseline glucose and HbA1c values will be used as covariates in analyses.

Such reporting will be done for all hypoglycemic events as well as for events classified as severe by the investigator and events classified by the investigator as SAEs.

### 7.4.9 Autoimmunity

Autoimmunity reactions will be described including severity and seriousness, as well as occurrence at ANA antibodies by the end of the study.

### 7.4.10 Injection site reactions

Events of injection site reactions including including SAEs and those reactions leading to drug discontinuation will be summarized by treatment group and presenting both the frequency and the severity of the events.

### 7.4.11 Vertigo/dizziness

Events of vertigo and dizziness will be summarized by treatment group and presenting both the frequency and the severity of the events.

### 7.4.12 Other safety topics

The following safety topics will be identified using a list of search criteria in the CRS unless specified otherwise.

- Routine risk of QT prolongation will be evaluated as described in section 7.3.4.
- Increased uric acid levels will be assessed by the laboratory analyses described in section 7.3.2.
- Long term effects on on kidney function will be assessed by the laboratory analyses described in section 7.3.2.
- Interactions with vaccines will be assessed separately by drug interactions.

## 8 Interim analyses

Interim analyses of efficacy, futility and safety will be carried out during the study. Two interim analyses of efficacy will be performed when about 50% of the target number of primary cardiovascular events have been accumulated and the second one when 75% of the planned number of events are available. Criteria for the interim and final analyses will be determined using a fixed Bonferroni split of the alpha allocated to the interim analyses and to the final analyses in order to protect the overall one-sided familywise type I error rate across

all analyses at 2.5%. The fixed total one-sided alpha allocated to both interim analyses of efficacy combined is 0.05%. Of this 0.01% is allocated to the first efficacy interim analysis and 0.04% allocated to the second efficacy interim analysis. The one-sided significance level for the final analysis is thus 2.45%.

Interim analyses for futility will be conducted simultaneously with the two analyses of efficacy. It should be noted that the efficacy criteria are not modified to "buy back" alpha based upon the presence of futility boundaries; this conservative approach ensures that the familywise type I error rate of the study is protected.

Full details on boundaries and stopping rules will be pre-specified in the Charter of the Data Monitoring Committee (DMC). Timing and number of safety analyses will also be specified in DMC charter.

Protocol amendment 8 introduced an additional early futility analysis to be performed at approximately 30% of the target number of 1400 patients have experienced CEC confirmed MACE. The analysis plan of the early futility was discussed in a separate document. No multiplicity adjustment will be made for the additional early futility analysis, because it will not result in rejecting a null hypothesis for efficacy superiority.

Interim analysis will be performed by an independent academic statistical data analysis center at [redacted] external to Novartis, who will not be involved in the trial conduct. The results will be reviewed by the Data Monitoring Committee (DMC).

Investigators, Novartis employees and others who are involved in the conduct of the trial and in the analysis of the final trial results, or who have contact with study centers, will remain blinded to the treatment codes and interim analysis results until all monitoring decisions have been made and the database has been locked for final analysis.

### DMC

In CACZ885M2301, a DMC will monitor the trial's progress, safety and efficacy on a regular basis or per the request of the DMC. Details regarding the reasons for the DMC and the grounds for stopping/continuing trials, as well as the DMC procedures are provided in the DMC charter.

## 9 Sample size and power considerations

Sample size and power considerations are addressed in section 9.6 and 9.7 of the clinical trial protocol.

## Appendix 1: Post-stroke functional assessment sub-study

Full details of the planned analyses for the Stroke Functional Assessment Sub-study, will be specified separately, prior to completion of the first Interim analysis.

In line with the primary objective of the post-stroke functional assessment sub-study to evaluate whether canakinumab arms facilitate functional recovery from stroke, the primary endpoint for this sub-study is the modified Rankin scale assessment 90 days post-stroke.

[redacted]

Note that while patients were initially randomly assigned to treatments, the subset of patients experiencing a stroke may represent a biased subset. The impact of the occurrence of fatal strokes in the main study on the results of the sub-study also needs to be considered.

Besides basic summary statistics the main study report will report an intention-to-treat analysis that will rank patients from best to worse as

- alive at trial end and no stroke,
- alive at trial end and stroke with best Rankin scale [redacted]
- alive at trial end and stroke with worst Rankin scale [redacted]
- dead at trial end

Ranking within these categories will be done with earlier stroke/death being worse than a stroke/death occurring later.

# 10 References

1. Allison PD (1995) Survival analysis using the SAS® System: A Practical Approach, Cary, NC: SAS Institute Inc.
2. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society, B, 57, 289–300.
3. Bhattaram VA, Siddiqui O, Kapcala LP, Gobburu JVS (2009) Endpoints and Analyses to Discern Disease-Modifying Drug Effects in Early Parkinson's Disease. The AAPS Journal 2009; Vol. 11, No. 3:456-464.
4. Bland JM, Altman DG (2004) Statistics Notes – The logrank test. BMJ 2004;328:1073.
5. Brannath W, Bretz F (2010) Shortcuts for locally consonant closed test procedures. J American Statistical Association ; 105(490),: 660-9.
6. Breslow NE (1974) Covariance analysis of censored survival data. Biometrics, 30:89-99.
7. Bretz F, Maurer W, Brannath W, et al (2009 A graphical approach to sequentially rejective multiple test procedures. Statistics in Medicine; 28(4): 586-604.
8. Brown H, Prescott R (2006) Applied mixed models in medicine. 2nd ed. 2006: John Wiley and Sons, Ltd, Chichester.
9. Buchanan TA (2007) (How) Can We Prevent Type 2 Diabetes? Diabetes 56:1502-1507.
10. Cox DR (1972) Regression models and life-tables (with discussion). J. R. Statist. Soc, B; 34:187-220.
11. Cox DR (1977) The role of significance tests. Scand J Statist 4: 49-70.
12. Cox and Snell (1989) The analysis of binary data. 2nd ed. 1989: Chapman and Hall, London.
13. Crowe BJ, Xia HA, Berlin JA, Watson DJ, Shi H, Lin SL, Kuebler J, Schriver RC, Santanello NC, Rochester G, Porter JB, Oster M, Mehrotra DV, Li Z, King EC, Harpur ES, Hall DB (2009) Recommendations for safety planning, data collection, evaluation and

reporting during drug, biologic and vaccine development: a report of the safety planning, evaluation, and reporting team. Clin Trials. 6(5):430-40.

14. Diabetes Prevention Program Research Group (2002) Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. NEJM; 346:393-403.
15. Dunnett CW (1955) A multiple comparison procedure for comparing several treatments with a control. J American Statistical Association; 50:1096-1121.
16. Efron B, Tibshirani RJ (1993) An Introduction to the Bootstrap. Chapman & Hall/CRC.
17. Ezetimibe Report on Deliberation results; March 6, 2007 PMDA Date of application October 31, 2003
18. EuroQoL Group (2009) EQ-5D User guide. Basic information on how to use EQ-5D, version 2.0, March 2009.
19. Fairclough D (2002) Design and analysis of quality of life studies in clinical trials. Chapman and Hall, New York.
20. Gail MH, Byar DP (1986) Variance Calculations for direct adjusted survival curves with applications to testing for no treatment effect. Biom J 28 (1986) 5, 587-599.
21. Genz A, Bretz F (2009). Computation of multivariate normal and t probabilities. Lecture Notes in Statistics, Vol. 195. Springer, Heidelberg.
22. Glynn RJ and Buring JE (1996) Ways of measuring rates of recurrent events. BMJ 1996; 312:364-7.
23. Glynn RJ, Stukel TA, Sharp SM, Bubolz TA, Freeman JL, Fisher ES (1993) Estimating the Variance of Standardized Rates of Recurrent Events, with Application to Hospitalizations among the Elderly in New England. Am J Epidemiol 1993;137:776-86.
24. Graham JW, Olchowski AE, Gilreath TD (2007). How Many Imputations are Really Needed? Some Practical Clarifications of Multiple Imputation Theory. Prev Sci (2007) 8:206–213.
25. Hakulinen (1982) Cancer survival corrected for heterogeneity in patient withdrawal. Biometrics, 38:933-942.
26. Holm S (1979) A simple sequentially rejective multiple test procedure. Scan J Statist 6:65–70.
27. Holzhauer B (in press) Design and Analysis of Diabetes Prevention Trials for Glucose-lowering Drugs. Statistics in Biopharmaceutical Research. DOI:10.1080/19466315.2013.861766
28. Kenward MG, Carpenter J (2007). Multiple imputation: current perspectives. Stat Methods Med Res 2007 16: 199.
29. Kenward MG, Molgenberghs G (1998) Likelihood based frequentist inference when data are missing at random. Statistical Science, Vol. 13, No. 3: 236-247.
30. Kenward MG, Roger JH (1997) Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. Biometrics, 53, 983–997.
31. Kvaløy and Neef (2004) Tests for the Proportional Intensity Assumption Based on the Score Process. Lifetime Data Analysis, 10, 139–157.
32. Levey AS, Coresh J, Greene T, Marsh J, Stevens LA, Kusek JW, Van Lente F for Chronic Kidney Disease Epidemiology Collaboration (2007). Expressing the Modification of Diet

in Renal Disease Study Equation for Estimating Glomerular Filtration Rate with Standardized Serum Creatinine Values. Clinical Chemistry 53:4: 766–772.

33. Little RJA, Rubin DB. Statistical Analysis with Missing Data. J. Wiley & Sons, New York, 1987.
34. Liu G, Gould AL (2002) Comparison of alternative strategies for analysis of longitudinal trials with dropouts. Journal of Biopharmaceutical Statistics 2002; 12: 2, 207 - 226.
35. Lunn M, McNeil D (1995) Applying Cox Regression to Competing Risks. Biometrics 51, 524-532.
36. Mallinckrodt CH, Kaiser CI, Watkino JG, Molenberghs G, Carroll RI. The effect of correlation structure on treatment contrasts estimated from incomplete clinical trial data with likelihood-based repeated measures compared with last observation carried forward ANOVA. Clinical Trials 2004; 1: 477-489, DOI: 10.1191/1740774504cn049oa.
37. Mallinckrodt CH, Watkin JG, Molenberghs G, Carroll RJ (2004). Choice of the primary analysis in longitudinal clinical trials. Pharmaceut. Statist. 2004; 3: 161–169, DOI:10.1002/pst.124.
38. Mallinckrodt CH, Kaiser CJ, Watkin JG, Detke MJ, Molenberghs G, Carroll RJ (2) Type I error rates from likelihood-based repeated measures analyses of incomplete longitudinal data. Pharmaceut. Statist. 2004; 3: 171–186, DOI:10.1002/pst.131.
39. Maruyama T., Takeda M et al Comparison of preventive effect on CV events with different statins. the Circle study Circ J 2011; 75 (8): 1951-9
40. Murray GD, Findlay JG (1988). Correcting for the bias caused by drop-outs in hypertension trials. Statistics in Medicine Volume 7, Issue 9: 941–946, DOI: 10.1002/sim.4780070905.
41. National Drug Monograph January 2012: www.pbm.va.gov
42. Nielsen B (1997) Expected Survival in the Cox Model. Scan J Statist 24: 275 – 287.
43. Peto R (1972) Discussion on Professor Cox's Paper. J. R. Statist. Soc., B; 34:205-207.
44. Pocock J, Clayton TC, Altman DG (2002) Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls. Lancet 2002; 359; 1686-89.
45. Rubin DB (1976). Inference and missing data. Biometrika 63, 581-592.
46. Rubin BR (1996). Multiple imputation after 18+ years. JASA; 91:473-489.
47. Saku K, Patrol study comparison of Rosuvastatin, Lovastatin and Pitavastatin:, Circulation Journal 2011;75 (6):1493-505
48. Senn S (1995) A personal view of some controversies in allocating treatments to patients in clinical trials. Statistics in Medicine, Vol. 14, 2661-2674.
49. Senn S (2008) Lessons from TGN1412 and TARGET: implications for observational studies and meta-analysis. Pharmaceut. Statist. 2008; 7: 294–301.
50. Schafer, JL (1997). Analysis of Incomplete Multivariate Data, New York: Chapman and Hall.
51. Siddiqui O, Hung HMJ O'Neill R(2009) MMRM vs. LOCF: A Comprehensive Comparison Based on Simulation Study and 25 NDA Datasets. Journal of Biopharmaceutical Statistics, 19: 2, 227-246, DOI: 10.1080/10543400802609797.
52. Stewart WH, Ruberg SJ (2000) Detecting dose response with contrasts. Statist. Med. 2000; 19:913-921.

53. Teramoto T. The clinical impact of pitavastatin: comparative studies with other statins on LDL-C and HDL-C
54. Therneau TM, Grambsch PM (2000) Modeling Survival Data – Extending the Cox Model. Springer.
55. Thomsen BL, Keidin N, Altman DG (1992) Letter to the editor: A note on the calculation of expected survival illustrated by survival of liver transplant patients. Statistics in Medicine, Vol. 11, 1527-1532.
56. Zhang X, Loberiza FR, Klein JP, Zhang M-J (2007) A SAS macro for estimation of direct adjusted survival curves based on a stratified Cox regression model. Computer methods and programs in biomedicine 88 (2007) 95-101.